



DISTRIBUTION STATEMENT A

Approved for public release
Distribution Unrestricted

NOWCASTING THUNDERSTORMS AT CAPE
CANAVERAL, FLORIDA, USING AN IMPROVED
NEUMANN-PFEFFER THUNDERSTORM INDEX

THESIS

Cindy L. Howell, Second Lieutenant, USAF

AFIT GM ENP 98M-05

DTIC QUALITY INSPECTED 4

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY
AIR FORCE INSTITUTE OF TECHNOLOGY

19980408121

Wright-Patterson Air Force Base, Ohio

AFIT/GM/ENP/98M-05

NOWCASTING THUNDERSTORMS AT CAPE
CANAVERAL, FLORIDA, USING AN IMPROVED
NEUMANN-PFEFFER THUNDERSTORM INDEX

THESIS

Cindy L. Howell, Second Lieutenant, USAF

AFIT/GM/ENP/98M-05

Approved for public release; distribution unlimited

The views expressed in this thesis are those of the author and do not reflect the policy or position of the Department of Defense or the United States Government.

**NOW-CASTING THUNDERSTORMS AT CAPE CANAVERAL, FLORIDA,
USING AN IMPROVED NEUMANN-PFEFFER THUNDERSTORM INDEX**

THESIS

Presented to the Faculty of the Graduate School of Engineering
of the Air Force Institute of Technology
Air University
Air Education and Training Command
In Partial Fulfillment of the Requirements for the
Degree of Master of Science in Meteorology

**Cindy L. Howell
2LT, United States Air Force**

March 1998

Approved for public release; distribution unlimited

AFIT/ENP/GM98M-05

NOW-CASTING THUNDERSTORMS AT CAPE CANAVERAL, FLORIDA, USING
AN IMPROVED NEUMANN-PFEFFER THUNDERSTORM INDEX

Cindy L. Howell
2 LT, United States Air Force

Approved:

Michael K. Walters
Lt Col Michael K. Walters
Chairman, Advisory Committee

4 MAR 98
Date

Cecilia A. Askue
Lt Col Cecilia Askue
Member, Advisory Committee

4 March 1998
Date

Dan E. Reynolds
Professor Dan E. Reynolds
Member, Advisory Committee

4 March 1998
Date

David Weeks
Dr. David Weeks
Member, Advisory Committee

4 March 98
Date

Acknowledgements

First and foremost, I would like to thank my thesis advisor, Lieutenant Colonel Michael K. Walters, without whose guidance this thesis would not have been nearly as complete or scientifically sound. I would also like to thank my committee members, Lieutenant Colonel Cecilia Askue, Professor Daniel E. Reynolds, and Dr. David Weeks, for their help in understanding various aspects of meteorology, statistics, and physics, respectively. Although Major Cliff Dungey was not a member of my committee, he was also a source of guidance for me, and that guidance is appreciated.

I would also like to acknowledge the role that Jason Tuell played in this endeavor. Dr. Tuell, former instructor of synoptic meteorology at the Air Force Institute of Technology, gave me the “applied meteorology” background that was so important in this project.

Next, I must thank Master Sergeant Pete Rahe, the Weather Lab Technician at the Air Force Institute of Technology. His tireless efforts to keep the lab computers operating correctly and to make back-up copies of my data are greatly appreciated.

I would also like to thank my sponsor, the 45th Weather Squadron at Patrick Air Force Base, Florida, and Mr. William Roeder, my point of contact. Mr. Roeder provided an insight into the operational issues that I was concerned about. He proved to be an invaluable resource throughout the course of this research project.

Also, I would like to thank my point of contact at the Air Force Combat Climatology Center (AFCCC), Captain Allen Rabayda. He quickly and efficiently ensured that I received the necessary data to complete my research.

Finally, I must acknowledge my appreciation to my friends for their unending patience and support. Especially, though, I must express my deepest appreciation to my family in Alabama. Even across the miles, they were my constant source of encouragement and support.

Cindy L. Howell

Table of Contents

Acknowledgments	iii
List of Figures	ix
List of Tables	x
Abstract	xi
1. Introduction	1
1.1 Significance of Problem	1
1.2 Background	2
1.3 Problem Statement	3
1.4 Benefit from Solving the Problem	3
1.5 Algorithm Tested	4
1.6 General Research Approach	4
1.7 Summary of Key Results	5
1.8 Thesis Organization	5
2. Literature Review	7
2.1 Basic Thunderstorm Theory	7
2.2 Previous Work	8
2.2.1 Neumann, 1968	9

2.2.2 Neumann, 1970	12
2.2.3 Neumann, 1971	14
3. Research Methodology	19
3.1 Introduction	19
3.2 Data Used	19
3.3 How Accurate is a Radiosonde?	21
3.4 Quality Control (QC) of the Data Set	24
3.5 Variables Included in the Algorithm	25
3.5.1 Climatological Probability of Thunderstorms	25
3.5.1.1 QC of the Climatological Probability of Thunderstorms	26
3.5.2 U and V Components of the 850-mb and 500-mb Winds	27
3.5.2.1 QC of the 850-mb and 500-mb Winds	28
3.5.3 Mean Relative Humidity from 600-800 mb	28
3.5.3.1 QC of the 600-800 mb Mean RH	29
3.5.4 Showalter Stability Index	29
3.5.4.1 QC of the SSI	31
3.6 Research Approach	31
4. Statistical Analysis	37

4.1 Introduction	37
4.2 The 2 X 2 Contingency Table	37
4.3 Measures of Accuracy	38
4.3.1 Hit Rate	38
4.3.2 False Alarm Rate	39
4.3.3 Probability of Detection	39
4.3.4 Threat Score	40
4.3.5 Skill Score	40
4.3.6 Bias Ratio	42
4.4 Pearson Correlation Coefficient	42
4.5 Test for Significance in the 2 X 2 Contingency Table	43
5. Results, Conclusions, and Recommendations	45
5.1 Introduction	45
5.2 Neumann's Forecast Verification	45
5.3 Using 24-Hour Persistence for Forecasting a Thunderstorm	46
5.4 Using 35% as Cutoff for Forecasting a Thunderstorm	48
5.5 Using 40% as Cutoff for Forecasting a Thunderstorm	49
5.6 Using 45% as Cutoff for Forecasting a Thunderstorm	50

5.7 Using 50% as Cutoff for Forecasting a Thunderstorm	52
5.8 Using 55% as Cutoff for Forecasting a Thunderstorm	53
5.9 Conclusions	55
5.10 Recommendations	56
5.11 Suggestions for Future Research	56
Appendix A: Wind Rose Plots for all Months	58
Appendix B: Name Changes and Map of Cape Canaveral	68
Appendix C: Input Constants for NPTI FORTRAN code	70
Appendix D: FORTRAN-77 Program for Calculating SSI	72
Appendix E: Polynomial Equations for all Months	76
Appendix F: P-Values and Chi-Squared Values at all Cutoff Levels	81
Appendix G: Pearson Correlation Coefficients	83
Appendix H: Example of a 2 X 2 Contingency Table	84
Appendix I: Histograms of Statistics at all Cutoff Levels	85
Bibliography	90
Vita	93

List of Figures

Figure 1: 15-Day Moving Averages (this study)	26
Figure A-1: 500-mb wind roses for May	58
Figure A-2: 850-mb wind roses for May	59
Figure A-3: 500-mb wind roses for June	60
Figure A-4: 850-mb wind roses for June	61
Figure A-5: 500-mb wind roses for July	62
Figure A-6: 850-mb wind roses for July	63
Figure A-7: 500-mb wind roses for August	64
Figure A-8: 850-mb wind roses for August	65
Figure A-9: 500-mb wind roses for September	66
Figure A-10: 850-mb wind roses for September	67
Map of Cape Canaveral area	69
Histogram of Statistics at 35% Cutoff Level	85
Histogram of Statistics at 40% Cutoff Level	86
Histogram of Statistics at 45% Cutoff Level	87
Histogram of Statistics at 50% Cutoff Level	88
Histogram of Statistics at 55% Cutoff Level	89

List of Tables

Table 1: Station IDs for Cape Canaveral	20
Table 2: Data used in this study	21
Table 3: Data used in Neumann's study	21
Table 4: SSI values and their operational definitions	30
Table 5: Verification of current NPTI using dependent data set	46
Table 6: Using persistence as a tool for forecasting thunderstorms	47
Table 7: Statistics for current and upgraded NPTI at 35% cutoff	49
Table 8: Statistics for current and upgraded NPTI at 40% cutoff	51
Table 9: Statistics for current and upgraded NPTI at 45% cutoff	52
Table 10: Statistics for current and upgraded NPTI at 50% cutoff	54
Table 11: Statistics for current and upgraded NPTI at 55% cutoff	55
Station IDs and geographic changes of Cape Canaveral	68
Input constants for current NPTI	70
Input constants for upgraded NPTI	71
P-values and chi-squared values	81
Pearson Correlation Coefficients	83

Abstract

Accurate thunderstorm forecasting is essential to the United States Air Force, especially the space program. The Neumann-Pfeffer Thunderstorm Index (NPTI) was introduced to forecast the probability of afternoon thunderstorms at Cape Canaveral, Florida, during the convective season. Very little further work has been done on the NPTI since its development in the 1960s. This thesis focuses on the NPTI, currently employed by the 45th Weather Squadron at Patrick AFB, Florida, and examines whether or not incorporating more data (15 years as opposed to 13 years) would significantly improve the NPTI.

All available upper air data and surface observations for Cape Canaveral were obtained from the Air Force Combat Climatology Center (AFCCC). Using the climatological probability of thunderstorms, the u and v components of the 850-mb and 500-mb winds, the 600-800 mb mean relative humidity, and the Showalter Stability Index (SSI), two linear regressions were performed, and probability equations were derived for May through September. Various statistical measures of accuracy were calculated in order to compare the current NPTI and the upgraded NPTI. Persistence forecasting was also considered.

Five cutoff percentages for forecasting a thunderstorm were considered. For most cutoff levels, the current NPTI and the upgraded NPTI were not significantly different; however, they were both better than forecasting persistence. For the higher cutoff

percentages, the upgraded NPTI showed slight improvement over the current NPTI. Interestingly, persistence forecasting produced better results than either NPTI at higher cutoff percentages. Hence, either more research is needed to improve the NPTI or a new algorithm should be developed to better forecast thunderstorms. Because the upgraded NPTI was shown not to be significantly different from the current NPTI, the current NPTI should continue to be used in operational thunderstorm forecasting until a more accurate algorithm is developed.

NOW-CASTING THUNDERSTORMS AT CAPE CANAVERAL, FLORIDA, USING
AN IMPROVED NEUMANN-PFEFFER THUNDERSTORM INDEX

1. Introduction

1.1 Significance of the Problem

Accurate thunderstorm forecasting is of utmost importance to the United States Air Force since thunderstorms directly impact the space program, as well as regular outdoor support activities (Manobianco et al., 1996: 654). The Neumann-Pfeffer Thunderstorm Index (NPTI) is used daily by the 45th Weather Squadron (WS) at Patrick Air Force Base during the convective season (May through September) to forecast the probability of afternoon thunderstorms for Cape Canaveral, Florida. It should be noted that when the NPTI was developed, the space shuttle launch site was known as Cape Kennedy. The site has changed names several times throughout the years. Appendix B includes a list of the various names as well as a map of the area. All future references in this paper will be made to Cape Canaveral, as the site is now called.

The 45th WS provides vital weather support for over 5000 pre-launch operations every year. The 45th WS is also responsible for protecting over seven billion dollars in resources and more than 25,000 people. Clearly, then, accurate thunderstorm forecasting is not only beneficial, but also vital to the mission of the 45th WS (Roeder, 1997).

1.2 Background

The NPTI is used to estimate the probability of afternoon thunderstorms at Cape Canaveral, Florida during the convective season. The algorithm, which was developed in the 1960s by Charles J. Neumann, uses temperature, dewpoint, relative humidity (RH), wind speed and direction, and the Showalter Stability Index (SSI) (Neumann, 1971: 7). Each of these parameters is obtained using data recorded from the morning radiosonde. The usual launch time for the morning radiosonde is 12Z. However, during times of strong convective activity or space shuttle activity, multiple balloons are launched. In the future context of this paper, the morning launch will refer to all radiosondes launched between 9Z and 15Z.

The five input variables used in the NPTI are climatological probability of thunderstorms, u and v components of the 850-mb winds, u and v components of the 500-mb winds, 600mb-800mb mean RH, and the SSI. The climatological probability of thunderstorms is based on a fifteen-day moving average of daily thunderstorm probability calculated from climatology (Neumann, 1968: 6). The 850-mb and 500-mb winds were shown by Neumann to be significant in terms of both speed and direction on days when thunderstorms occurred. His study also concluded that the 600-800 mb layer is the most important layer for the presence of moisture on thunderstorm days (Neumann, 1971: 7). Finally, the SSI is one of the many tools used in meteorology to examine the stability of the atmosphere; this stability is then used to determine the potential for severe weather such as thunderstorms (AWS/TR-79/006: 5-35). From a possibility of over 250

predictors, Neumann determined that these five explained the most variance in thunderstorm occurrence in his study (Neumann, 1971: 7). The NPTI is discussed in greater detail in Chapter 2.

1.3 Problem Statement

In his 1968 study, Neumann used only 13 years of data. Little or no work has been done on the NPTI since then. The current NPTI, which is based on only a few years of data taken more than 30 years ago, could potentially be improved by using a larger data set to recalculate the regression coefficients. This study examined whether or not the current NPTI can be improved by including a total of 15 years of data. The current NPTI, the upgraded NPTI, and persistence forecasting were compared by computing various statistical measures of accuracy. Finally, both NPTIs were validated using a two-year independent data set.

1.4 The Benefit from Solving the Problem

An estimated 30% of all space shuttle launches are either delayed or cancelled due to weather, specifically thunderstorms, and each time a launch is cancelled, or scrubbed, it costs an estimated one million dollars to de-fuel the shuttle and prepare it for its next potential launch (Roeder, 1997). Between the years of 1981 and 1994, nearly 75% of all space shuttle countdowns were delayed or scrubbed; almost half of these were due to weather (Hazen et al., 1995: 273). At present, 40% of all thunderstorm forecasts result in false alarms, and another 10% fail to provide the desired lead-time (Roeder, 1997). An improved NPTI would provide more accurate thunderstorm forecasting which would, in

turn, save the Air Force several valuable resources: equipment, finances, and, most importantly, human life.

1.5. Algorithm Tested

The Neumann-Pfeffer Thunderstorm Index was examined in this study. Neumann's study involved performing a multiple regression analysis on 13 years of data (Neumann, 1971: 2). By using similar multiple regression techniques and 15 years of data, the current NPTI was studied and the regression coefficients in the predictor equations were recalculated. Both NPTIs were validated using a two-year independent data set and were compared by computing various measures of accuracy. The methodology and results will be discussed in the subsequent chapters.

1.6 General Research Approach

This research project consisted of three main tasks: data collection and quality control, performing the regression, and interpretation and analysis of the regression. After obtaining the necessary data from the Air Force Combat Climatology Center (AFCCC), the data was put through several quality control checks, which will be discussed in Chapter 3. Then, the relevant data was extracted and manipulated into a more useable format using a series of FORTRAN-77 programs. In preparation for the second task, the data was split by month, thus yielding five one-month data sets (May through September). After splitting the data, it was imported into Statistix[©] by month and transformed following the method used by Neumann (Neumann, 1971: 9). Then task two, running the regression, was performed. Involved in this task were choosing an appropriate model in Statistix[©], running the regression model, and formatting the output

in tabular form. Finally, task three required an interpretation and analysis of the regression analysis.

1.7 Summary of Key Results

The current NPTI and the upgraded NPTI were validated against two independent years of data. Although each index yielded certain measures of accuracy that were a bit higher than the other index, these differences were generally insignificant. Furthermore, forecasting persistence did not usually produce better results than either NPTI. This was not the case at higher cutoff percentages, where persistence performed as well or better than either version of the NPTI. Because the differences in the two indices are so minuscule, the current NPTI should continue to be used operationally. However, it is concluded that a more accurate method of forecasting thunderstorms is needed.

1.8 Thesis Organization

Chapter 2 gives a discussion of Neumann's early work on the Neumann-Pfeffer Thunderstorm Index, as well as some background information and concepts that were vital to the understanding of the problem.

Chapter 3 presents an in-depth discussion of the research approach and techniques, outlined previously in this chapter.

Chapter 4 provides a statistical analysis of the multiple regression results. Included in this discussion are the 2 X 2 contingency table, hit rate (HR), false alarm rate (FAR), probability of detection (POD), threat score yes (TS-yes), threat score no (TS-no), skill score (SS) against persistence, and bias ratio (B). The Fisher-Irwin test, p-values, chi-squared values, and the Pearson correlation coefficient are also discussed.

Finally, Chapter 5 presents the key results gleaned from this study. Recommendations for operational use and suggestions for further research projects are also given.

2. Literature Review

2.1 Basic Thunderstorm Theory

Thunderstorms may form when three conditions are met. These conditions are low-level atmospheric moisture, lift, and atmospheric instability (Wallace and Hobbs, 1977: 86; McGinley, 1986: 669).

Low-level moisture is perhaps the most essential ingredient necessary for the development of thunderstorms. Two potential sources for this moisture, especially for the Florida peninsula, are the Atlantic Ocean and the Gulf of Mexico (Jessup, 1972: 654; Weiss, 1992: 964; McGinley, 1986: 669). Another possible source for moisture at Cape Canaveral is the vast river network that surrounds the area (Cetola, 1997: 2-4).

Lift, caused by converging air in the low levels of the atmosphere, is also crucial in thunderstorm development (Zhong and Takle, 1993: 1185; Zhong and Takle, 1992: 1426). At Cape Canaveral, this lift, or vertical motion, is usually associated with the sea breeze circulations (Wilson and Megenhardt, 1997: 1507; Cetola, 1997: 2). A sea breeze circulation forms when the land temperature is warmer than the adjacent water temperature. This usually occurs during the day when solar radiation heats the land more than it heats the water. This heating results in a shallow thermal low, or heat low, forming over the land and a shallow thermal high forming over the water. Because the wind blows from high pressure to low pressure, when the temperature difference between the land and the water is great enough, a sea breeze forms and moves ashore. At the leading edge of the sea breeze, referred to as the sea breeze front, there is an area of

enhanced low-level convergence and vertical motion, which is often associated with thunderstorm formation (Cetola, 1997: 2; McGinley, 670). Lift, however, is not always a result of the sea breeze front. Lift can also be caused by speed convergence or directional convergence. If the winds are blowing from the same direction and the tail wind is faster than the lead wind, the result is speed convergence. If, on the other hand, the winds are blowing from opposite directions toward a central region, regardless of wind speed, the result is direction convergence. In either case, low-level convergence and lift occur. A mechanism that can lead to greater instability is the position of the wind speed maximum at the level of the jet stream. If the left-front and right-rear quadrants of the jet max are positioned above regions of instability, lift is further enhanced (McGinley, 1986: 672).

The final ingredient necessary in the formation of thunderstorms is atmospheric instability (Weiss, 1992: 964). The main cause for instability in the atmosphere is surface heating (Zhong and Takle, 1992: 1437). When a parcel is warmer than its environment, it rises and continues to rise until its temperature becomes cooler than its environment. As the parcel is rising, it is unstable since its temperature is warmer than the temperature of the atmosphere (Wallace and Hobbs, 1977: 85).

Neumann also mentioned these criteria and added that Florida is a favored area for thunderstorm development since Florida possesses these conditions so often during the convective season (Neumann, 1968: 1).

2.2 Previous Work

The bulk of the background work for this project was done by Neumann in the 1960s. He studied the frequency and duration of thunderstorms at Cape Canaveral,

various characteristics of thunderstorms, and the properties that govern the atmosphere as it pertains to thunderstorm development. After much studying and testing, he introduced the Neumann-Pfeffer Thunderstorm Index, to improve the forecasting of probability of thunderstorms on a particular day based on data obtained from the morning radiosonde. From a pool of more than 250 potential predictors, he decided that five predictors were consistently significant on days during which a thunderstorm occurred. These five predictors, as previously discussed, are the u and v components of the 850-mb and 500-mb winds, the 600-800 mb mean relative humidity, the Showalter Stability Index, and the day number, which is a function of the climatological frequency of thunderstorms.

Although much research has been done to study thunderstorm forecasting and thunderstorm development in Florida, aside from the efforts of Neumann, no further work or revision has been done on the NPTI.

2.2.1 Neumann, 1968

Neumann mentioned three main reasons why Florida is one of the major regions of thunderstorm activity (Neumann, 1968: 1).

1. There is an abundant supply of low-level moisture along with the conditional instability needed to trigger thunderstorms.
2. The sea-breeze convergence over the Florida peninsula provides the lift mechanism required for thunderstorm development.
3. In some cases, the synoptic setting is such that thunderstorm activity is enhanced.

This assessment is in agreement with several noted authors who mention the ingredients necessary for the development of thunderstorms.

Neumann then went on to discuss the data he used in his analysis. The surface observations from the years 1951, 1952, and 1957-1967 were used to compute the climatological probability of thunderstorms. From 1953-1956, the surface observations were quite sparse; hence, Neumann did not use them.

He used the upper air observations, from the years 1957-1969. While the upper air data from the years 1950-1955 was available, it was determined to be less accurate than data from later years. During these years, the wind direction was reported to the nearest integer multiple of 22.5 degrees. Beginning in 1956, the wind was measured to the nearest degree. To clarify, before 1956, if the wind was actually blowing from 19 degrees, it was reported as blowing from 22.5 degrees. The same wind measurement after 1956 was reported as blowing from 19 degrees. When converting the winds to u and v components from data taken prior to 1956, error could occur, depending on the magnitude of the wind speed and direction. The discrepancies in the wind measurements are easily seen in wind roses plotted for each month. Appendix A includes wind roses plotted for each month using the wind data before 1956 and then from 1956 forward.

For this reason, the first year of upper air observations included in this study was 1957. Despite the lower quality of the upper air observations taken prior to 1956, the surface observations from 1950 and later, although sparse, were used where available in computing the climatological frequency of thunderstorms.

Neumann pointed out that the observation site for Cape Canaveral has changed several times throughout the years. A list of the different observation sites, as was noted

in Chapter 1, can be found in Appendix B. These slight geographical shifts, however, are insignificant (Neumann, 1968: 3).

Neumann described how he calculated the climatological frequency of thunderstorms using a 15-day moving average. By trial and error, other n-day moving averages were rejected either because of excessive data smoothing or because these other n-day averages were computationally expensive (Neumann, 1968: 6-7). In keeping with Neuman's method, the climatological probabilities in this study were also computed using 15-day moving averages.

After dividing the convective season into eight distinct periods and listing their characteristics, he noted five significant features of the thunderstorm pattern at Cape Canaveral (Neumann, 1968: 7-14).

1. There is a double peak in the seasonal thunderstorm cycle. The first peak occurs on June 30 and the second on August 3, on average.
2. Between early March and early April, there is a secondary maximum of thunderstorm activity.
3. The main convective season was identified as May 16 through September 22; on 25% of the days in this interval, thunderstorms can be expected.
4. From December 28 through January 12, there were no thunderstorms recorded over the 13 years Neumann used in his study. It should be noted that the present study focused only on the months from May through September, the convective season.
5. Most late night and early morning thunderstorms occur from mid-August through mid-September.

Neumann computed the probabilities for thunderstorm occurrence on a given day as well as the conditional probabilities for thunderstorm occurrence over an extended period (Neumann, 1968: 10-19). Conditional probabilities were not considered in Neumann's study or in the present study.

2.2.2 Neumann, 1970

As in his previous report, Neumann discussed the eight periods of the thunderstorm cycle at Cape Canaveral. Since Neumann discussed these eight periods in both of his reports, it seems worthwhile to list them below (Neumann, 1970: 5-6).

1. From November through early March, thunderstorms are usually the result of instability or convergence associated with synoptic-scale disturbances.
2. From early March through early April, there is a marked increase in thunderstorm activity mostly due to prefrontal squall lines.
3. Due to a sharp decrease in frontal activity, there is a slight decline in thunderstorm activity in mid-April.
4. From late April through June, when solar heating begins to increase, there is an increase in thunderstorm activity.
5. In the first half of July, there is a slight decline in activity; this can best be explained by the positioning of the mid-tropospheric (500-mb) ridge line.
6. For the same reason mentioned in period five, the latter half of July through early August also shows a decline in thunderstorm activity.
7. Due to a gradual decrease in solar heating, there is a corresponding decrease in thunderstorm activity from early August through the first third of September.

8. From the later two-thirds of September through October, there is a rapid decline in thunderstorm activity. This is a direct result of the decrease of solar radiation at that time of the year.

In this report, Neumann provided thunderstorm probabilities for Cape Canaveral based on three predictors: the 12 GMT 3000-ft wind direction, the 3000-ft wind speed, and the date (Neumann, 1970: 5). He used the same data that he used in his first study (Neumann, 1970: 9).

Neumann discussed the climatological characteristics of the speed and direction of the 3000-ft winds. He plotted a series of ellipses depicting the u and v components of the 3000-ft winds using the same 13-year period that he had used in his previous study. These ellipses show the relative magnitudes of the u and v components and provide a broad view of the wind field (Neumann, 1970: 9-14).

Neumann used the regression estimation of event probabilities (REEP) approach. When using this approach, it is possible, although unlikely, to obtain a probability outside the range from 0 to 1 (Wilkes, 1995: 183). Hence, the main purpose of the ellipses was to bound the u and v components so the binomial probability distribution yields only values between 0 and 1.

After examining the effects of wind speed alone and wind direction alone, he concluded that a combination of both speed and direction of the low-level wind was the single most important factor for thunderstorm occurrence (Neumann, 1970: 8, 20). Several different frequency and probability distributions of thunderstorm occurrence, wind speed, and wind direction were plotted to show the frequency of thunderstorms that

occurred when speed and direction were used as separate predictors (Neumann, 1970: 17, 19). Using his plot of thunderstorm probability based only on the 3000-ft wind direction, Neumann made several observations. First, from November through April, northeasterly winds never produced afternoon thunderstorms. Perhaps, for this reason, Neumann added a subjective correction factor in the NPTI. This is merely speculation, however, as Neumann never explained the reasoning behind this correction factor. The subjective correction factor will be discussed later in this chapter.

Next, Neumann noted that from early July through mid-August, west and southwest winds produced thunderstorms at least 75% of the time. Finally, during the early and late portions of the convective season, maximum thunderstorm activity occurs with south or southwest 3000-ft winds (Neumann, 1970: 22). Neumann stressed that day-to-day persistence should also be considered in operational thunderstorm forecasting. Using August 1 as an example, given that a thunderstorm occurred on the previous day, there is a 70% chance of having another thunderstorm on that day (Neumann, 1970: 29).

As he did in his previous work, Neumann computed probabilities for thunderstorm occurrence on a single day as well as conditional probabilities over various time periods and constructed probability tables to that end (Neumann, 1970: 33-63).

2.2.3 Neumann, 1971

Neumann began by explaining why accurate thunderstorm forecasting at Cape Canaveral is vital to the United States space program. These reasons were discussed in Chapter 1. After recapping his two previous thunderstorm studies at Cape Canaveral, Neumann introduced the idea of using multiple regression techniques to observe the

relationship between the five independent predictors and the dependent variable, whether or not a thunderstorm actually occurred. Following a binomial distribution, if a thunderstorm occurred, a 1 was assigned to that day, and a 0 was assigned if a thunderstorm did not occur. Next, Neumann showed plots of the 15-day moving averages as functions of thunderstorm frequency for various time intervals (Neumann, 1971: 1-3).

Neumann found nonlinear trends in the data to be statistically significant. Therefore, he used second and third-order polynomials to represent the five independent variables rather than the five predictors themselves. The general forms of the polynomial equations that were used to transform the variables into their nonlinear forms are shown below (Neumann, 1971: 9):

$$F(X1) = A_0 + A_1S + A_2T + A_3ST + A_4S^2 + A_5T^2 + A_6S^3 + A_7S^2T + A_8ST^2 + A_9T^3 \quad (1)$$

$$F(X2) = B_0 + B_1U + B_2V + B_3UV + B_4U^2 + B_5V^2 + B_6U^3 + B_7U^2V + B_8UV^2 + B_9V^3 \quad (2)$$

$$F(X3) = C_0 + C_1RH + C_2RH^2 + C_3RH^3 \quad (3)$$

$$F(X4) = D_0 + D_1 SSI + D_2 SSI^2 \quad (4)$$

$$F(X5) = E_0 + E_1 DAY + E_2 DAY^2 \quad (5)$$

Where:

S, T = u, v components of 850-mb wind in knots

U, V = u, v components of 500-mb wind in knots

RH = 600-800 mb mean relative humidity in percent

SSI = Showalter Stability Index in degrees Celsius

DAY = Day number

X1 = 850-mb wind in knots

X2 = 500-mb wind in knots

X3 = 600-800 mb mean relative humidity in percent

X4 = Showalter Stability Index in degrees Celsius

X5 = Day number

Once Neumann defined the five variables in this manner, he performed the first of two nonlinear multiple regressions; he regressed the combinations of variables on the right side of equations (1-4) against the set of 0s and 1s that represent the occurrence of a thunderstorm. He regressed thunderstorm frequency against day number in equation (5). Then, he extracted the coefficients for each term and substituted them as the constants in equations (1-5). The next step was to insert raw data (u and v components, RH, SSI, and

day number) into these five polynomials, evaluate the polynomials for each day, and regress the polynomials against the binomial distribution of 0s and 1s. From this second regression, the monthly prediction equations and regression coefficients for May through September were defined as follows (Neumann, 1971: 4-9):

$$P(\text{may}) = H_0 + H_1 F(X_1) + H_2 F(X_2) + H_3 F(X_3) + H_4 F(X_4) + H_5 F(X_5) \quad (6)$$

$$P(\text{jun}) = K_0 + K_1 F(X_1) + K_2 F(X_2) + K_3 F(X_3) + K_4 F(X_4) + K_5 F(X_5) \quad (7)$$

$$P(\text{jul}) = L_0 + L_1 F(X_1) + L_2 F(X_2) + L_3 F(X_3) + L_4 F(X_4) + L_5 F(X_5) \quad (8)$$

$$P(\text{aug}) = N_0 + N_1 F(X_1) + N_2 F(X_2) + N_3 F(X_3) + N_4 F(X_4) + N_5 F(X_5) \quad (9)$$

$$P(\text{sep}) = P_0 + P_1 F(X_1) + P_2 F(X_2) + P_3 F(X_3) + P_4 F(X_4) + P_5 F(X_5) \quad (10)$$

He noted that the importance of individual predictors differed from month to month. However, for the sake of uniformity, all five predictors were included in each month's prediction equation (Neumann, 1971: 7).

In order to run the NPTI using Neumann's FORTRAN code, the constants for each month's polynomials must be read into his program. A list of these constants, along with the constants derived in this study, is included in Appendix C. He incorporated a subjective correction into his code to account for days with strong easterly winds. As mentioned earlier in this chapter, the reason for this correction factor may have been because northeasterly winds never produced afternoon thunderstorms in his study. Although he never explicitly explained his reason for this, the present study tested the current and revised NPTI both with and without the correction factor.

When the NPTI is run, the result is a “yes-no” thunderstorm forecast. It should be noted that, operationally, different percentages for each month can be used as the cutoff for forecasting a thunderstorm.

Finally, Neumann briefly mentioned his verification process, in which he used the years from 1957-1969 as a dependent data set. He recorded the observed occurrence rate (obtained from the dependent data set) and the forecast probability of thunderstorm occurrence (obtained from running the NPTI). He found that “forecast probabilities of less than 0.50 are too high and those above 0.50 are too low” (Neumann, 1971: 26). However, forecasts near 0.50 were generally correct.

These results were obtained from the verification of the dependent data set, although the results were similar when 1970 was used as an independent data set (Neumann, 1971: 26, 30). Neumann was not sure of the reason for this bias, but he made it clear that more independent data should be used to more accurately assess the performance of the algorithm (Neumann, 1970: 26, 30).

3. Research Methodology

3.1 Introduction

It is important to understand precisely what data was used in this study and how accurate the data is. In order to understand the accuracy of the data, it is necessary to have a basic notion of how the measurements were taken, how accurate the measurements are, and what quality control checks were employed to ensure the data was quality data. This chapter presents a discussion of these topics. Finally, a thorough discussion of the research techniques employed in this project is given.

3.2 Data Used

The surface observations from 1950-1996 at the Cape Canaveral, Florida, observation site, were used to determine the climatological probability of thunderstorm occurrence; this process is described later in this chapter. The surface observations were also used to verify the NPTI and to build the 2 X 2 contingency tables, which were used to derive and compute various measures of accuracy such as HR, FAR, POD, TS-yes, TS-no, and SS, as well as bias ratio.

Upper air observations in the interval from 9Z-15Z were also vital in the completion of this project. Variables such as the 850-mb and 500-mb winds, the 600-800 mb mean RH, and the SSI were all calculated from information extracted from the upper air observations. Each upper air observation site is identified by a four-letter code, or station ID. Cape Canaveral's station ID has changed several times since 1950, as shown

in Table 1. In searching for both the surface observations and the upper air observations, these ID changes had to be considered.

Table 1. Station IDs for Cape Canaveral (Roeder, 1997)

Inclusive dates	Station IDs for Cape Canaveral
June 1950-16 March 1978	KXMR
17 March 1978-31 July 1980	KX68
11 February 1993-19 May 1993	KQCH
20 May 1993-16 June 1993	KKSC
17 June 1993-Present	KTTS

This study was originally designed to examine 30 years of data but ended up using only 17 years. While 46 years of surface observations were available, only 17 years of upper air data were recovered. Some of the other years had missing variables (identified by the string 999) in the data and could not be used. Other years had simply not been archived by AFCCC. Tables 2 and 3 list, by month, which years of surface observations and upper air observations were used in this study and in Neumann's study.

In order for a day to be included in the regression, all five variables had to be available for that particular day. In other words, the data set was reduced to only days that included all five predictors, and these days were then used to build the regression model. This process of matching days drastically depleted the data set.

Table 2. Data Used in This Study

	May	June	July	Aug	Sept
Surface Observations	1951-1953 1957-1996	1951-1996	1951-1953 1956-1996	1950-1953 1956-1996	1950-1996
Upper air Observations	1957-1969 1983,1985 1987,1988	1957-1969 1983,1985 1987,1988	1957-1969 1983,1985 1987,1988	1957-1969 1983,1985 1987,1988	1957-1969 1983,1985 1987,1988

Table 3. Data Used in Neumann's Study (Neumann, 1971: 2)

Observations	All months
Surface Observations	1951-1952 1957-1967
Upper air Observations	1957-1969

3.3 How Accurate is a Radiosonde?

The radiosonde that has been used by the United States for 30 years consists of a “temperature-compensated aneroid capsule that moves a lever arm across a commutator plate” (Golden, Serafin, Lally, and Facundo, 1986: 51). This design and the lever arm allow five times the deflection at 50 mb as at 1000mb. At the surface, this “baroswitch” is accurate to + or – 1 mb. Pressure measurements are accurate to + or – 2 mb near 500

mb, and to + or - 1 mb at 10 mb (Golden et al., 1986: 51). This type of design seems to be both accurate and reliable.

To measure temperature and humidity, a rod thermistor is used. Its diameter is approximately 0.7 mm, and it is 1 to 2 cm long. The thermistor is coated with a lead carbonate pigment; this type of coating helps to reduce solar heating, which can cause errors in temperature measurements. The rod has errors of 1-2 degrees Celsius above 25 km due to its high absorption in the infrared. Lag of the rod thermistor is another source of error. To that end, there is a correction factor that should be used for all radiosonde measurements (Golden et al., 1986: 51, 52):

$$T = MT + (LR)(AR)(LC) \quad (11)$$

Where:

T = actual temperature

MT = temperature measured by the radiosonde

LR = lapse rate

AR = ascent rate of the balloon

LC = lag constant of the thermistor

For humidity measurements, a carbon sensor made up of a "thin coating of a fibrous material on a glass or plastic substrate" is used (Golden et al., 1986: 52). The accuracy of this sensor is generally 5-7% in relative humidity for most temperatures.

The sensors the United States currently employs has a systematic bias of about 2-4% around saturation for temperatures above freezing. In 1985, humidity equations were re-derived to account for this bias (Golden et al., 1986: 52).

For the most part, the wind speed and direction in “synoptic-scale geostrophic flow pattern are representative” of the atmosphere (Golden et al., 1986: 52, 53). Large gradients are smoothed by various averaging techniques. As a result of these techniques, in areas near the jet stream or near a jet maximum, the wind measurements can be underestimated by as much as 20% (Golden et al., 1986: 53). However, generally, the wind measurements are accurate to within one meter per second (Roeder, 1998).

The first successful radio direction-finding system was the SCR-658, which was developed in World War II. The system operated at 400 MHz, and it used two operators to steer an antenna array to determine the direction of the radiosonde transmitter. At present, the United States uses a similar, but faster, design. The current system operates at 1680 MHz, and it uses an automatic tracking system. In order to determine the height of the radiosonde, pressure readings are converted, using the hydrostatic equation, into their equivalent altitudes. At 10 km, the error is generally 20 m, and at 30 km, the error can be as much as 100 m. The WBRT, the radiosonde used by the United States, uses the computed altitude and elevation angle to find the horizontal distance to the radiosonde. Due to the potentially high errors, many radiosonde launching stations also use a transponder attachment to measure slant range; this improves accuracy at low elevation angles.

3.4 Quality Control of the Data Set

The data was put through rigorous quality control checks before it was used in this study. First, the Air Force Global Weather Center, who submitted the data to AFCCC for archival, ran the data set through a series of 211 systematic algorithms. These algorithms, designed for use on the planetary scale, check the data for extreme measures; if extreme measures are detected, the algorithm corrects them, if possible (AFCCC/TN-96/001, 1996). Once AFCCC received the data, it was quality controlled again. A scatter plot of the data was constructed to look for outliers. Any outliers were flagged. Later, the flagged data were checked manually. The biggest potential problem for the flagged data is simply bad key entry. For example, a temperature of 10 degrees Celsius should be entered as “10.0”. An entry of “100”, where the decimal point is out of place, would be a bad key entry. Any such entries were manually corrected (Rabayda, 1998).

An initial quality control measure that was performed after the data was received from AFCCC included choosing a random sample of approximately 20% of the entire data set and confirming that the data was plausible. For instance, if the temperature at 500 mb was listed in the data set as being 65 degrees Celsius, the day would be flagged as “bad”; a temperature that high at 500 mb is virtually impossible. Only a few days were flagged as “bad”; these “bad” days were eliminated from the study. Another preliminary quality control check was to subtract the dewpoint from the temperature. Because the dewpoint can never be higher than the temperature, a negative difference could be indicative of erratic data; in this case, too, the day was flagged as “bad” and the day was

eliminated from consideration. All of the data was checked for negative differences.

Besides the initial quality control measures, other measures were employed to quality control the calculations of the five input variables of the NPTI. A description of these other quality control checks follows as each individual variable is discussed.

3.5 Variables Included in the Algorithm

Five variables, or predictors, are used in the NPTI algorithm. These are the climatological probability of thunderstorms, the u and v components of the 850-mb and 500-mb winds, the 600-800 mb mean relative humidity, and the Showalter Stability Index. Each of these predictors is examined below.

3.5.1 Climatological Probability of Thunderstorms

The climatological probability, or frequency, of thunderstorms was computed from 46 years of surface observations taken at the Cape Canaveral observation site. The events of either a thunderstorm or thunder heard were tallied for each day of the convective season, and a probability was computed for each day and smoothed using a 15-day moving average (Neumann, 1968: 6). The general form of the formula is as follows:

$$FREQ(Day_N) = \sum_{I=N-7}^{N+7} \frac{Total(N)}{15K} \quad (12)$$

Where:

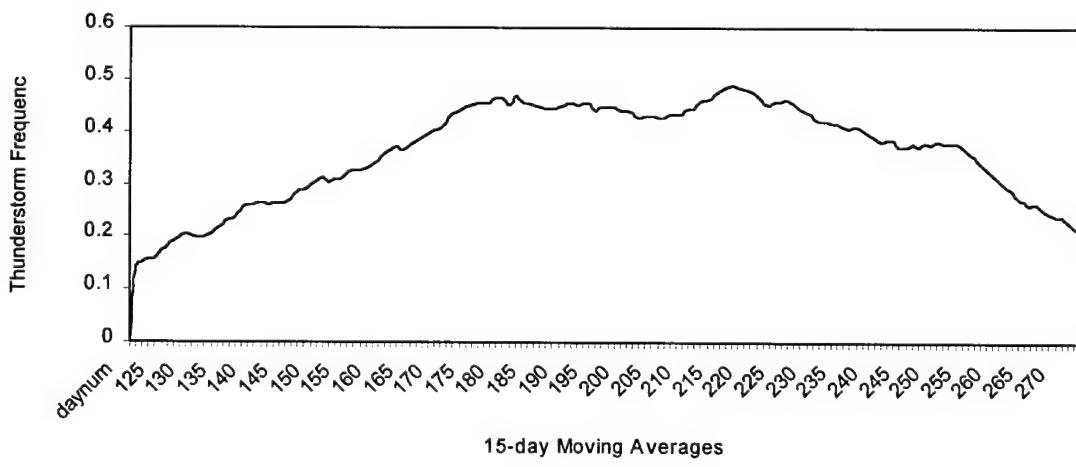
Total (N) = total number of occurrences of either thunderstorm or thunder heard over the K years on day N.

K = number of years of surface data used for each day (K was a constant for the days in each month, although K did vary from month to month.)

3.5.1.1 Quality Control of the Climatological Probability of Thunderstorms

The calculated fifteen-day moving averages and the climatological frequencies of thunderstorms from this study were compared to those of Neumann. The results were comparable. Several features of the two plots should be pointed out. First, there are distinct double peaks in the frequencies around late June-early July (day numbers 178-184) and in early August (day numbers 213-218). The minimum falls near the latter third of July (day numbers 197-204). These characteristics, discussed by Neumann, were mentioned in Chapter 2 (Neumann, 1968: 7-14). A plot of the fifteen-day moving averages against the frequencies of thunderstorms from this study is shown in Figure 1.

Figure 1. 15-day moving averages and thunderstorm frequencies (this study)



3.5.2 U and V Components of the 850-mb and 500-mb Winds

It is common practice in meteorology to separate the wind speed and direction into orthogonal x and y components, referred to as u and v. Trigonometric functions are used to convert the speed and direction into u and v components. These formulas are shown below (Neuman, 1968: 40):

$$U8 = \sin[(Dir8)(0.0174533) + \pi]Spd8 \quad (13)$$

$$V8 = \cos[(Dir8)(0.0174533) + \pi]Spd8 \quad (14)$$

$$U5 = \sin[(Dir5)(0.0174533) + \pi]Spd5 \quad (15)$$

$$V5 = \cos[(Dir5)(0.0174533) + \pi]Spd5 \quad (16)$$

Where:

U8, U5 = u component at 850 mb and 500 mb

V8, V5 = v component at 850 mb and 500 mb

Dir8, Dir5 = wind direction at 850 mb and 500 mb, in degrees

Spd8, Spd5 = wind speed at 850 mb and 500 mb in knots

3.5.2.1 Quality Control of the 850-mb and 500-mb Winds

The u and v components of the 850-mb and the 500-mb winds were also checked for quality. As was mentioned above, the few days with unusually high or low wind speeds were eliminated in the initial quality control check. Wind roses were plotted for the wind speed and direction for the years prior to 1956 and for the years starting with 1956. These wind roses, which can be found in Appendix A, plainly illustrate the different methods, discussed earlier, that were used to record wind direction prior to 1956. After studying these plots, it was concluded that this is the reason Neumann excluded the years before 1956 from his study.

To ensure that the computed values of u and v were correct, a random 10% of the data were plotted by hand. The wind speeds and directions were plotted in a Cartesian coordinate system using the trigonometric identities discussed later in this chapter. These hand-plots seemed to match quite well with the computed u and v components.

3.5.3 Mean Relative Humidity from 600-800 mb

The layer from 600-800 mb was shown by Neumann to be the most significant layer for the presence of moisture in relation to the occurrence of thunderstorms because the layer displayed the highest correlation between moisture and thunderstorm occurrence (Neumann, 1968: 7). To compute the mean relative humidity of a layer, each level was weighted logarithmically to account for atmospheric pressure being non-linear. The formula, adapted from the Air Weather Service's Technical Report 83/001, follows below:

$$MeanRH = \frac{1}{\ln(800) - \ln(600)} \times \sum_{I=1}^4 [0.5(RH(I) + RH(I+1)) \times \ln(P(I)) - \ln(P(I+1)))] \quad (17)$$

Where:

RH (I) is the relative humidity at level I=1

P (I) is the pressure at level I=1, 800 mb in this case

3.5.3.1 Quality Control of the 600-800 mb Mean RH

Next, the 600-800 mb mean RH was quality controlled. Once again, the original data set was checked for any value of RH that exceeded 100, although none was found. The weighted RH calculation, described by the equation (17) above, was compared to an arithmetic average of the RH values over a random 20% of the entire data set. To obtain the arithmetic average, the RH values at 600-mb, 650-mb, 700-mb, 750-mb, and 800-mb were added; then the sum was divided by five. The arithmetic average was compared to the weighted RH to ensure that the algorithm was calculating it correctly.

3.5.4 Showalter Stability Index

The Showalter Stability Index is often used to determine whether or not thunderstorms are likely to occur, and if so, their potential severity. To compute the SSI manually requires several steps. Given a Skew T, Log P sounding, a line is drawn dry adiabatically from the 850-mb temperature to the lifted condensation level, or LCL.

From the LCL, a line is drawn along the saturated adiabat until it intersects 500 mb. The temperature at this point is called T' . Next, algebraically subtract T' from T , the actual temperature at 500 mb. The remainder is the SSI.

In this study, the SSI was calculated using a FORTRAN-77 program. This program is included in Appendix D.

A SSI value of less than +3 means that showers are probable and some thunderstorms could occur. A value in the range of +1 to -2 indicates a marked increase for potential thunderstorm activity, while a value of less than -3 is usually associated with severe thunderstorms (AWS/TR-79/006). These values are summarized in Table 4.

Table 4. SSI Values and Their Operational Definitions

SSI value	Definition
Less than +3	Showers probable; thunderstorms possible
Between +1 and -2	Marked increase for potential thunderstorm activity
Less than -3	Associated with severe Thunderstorms

3.5.4.1 Quality Control of the SSI

Finally, quality control measures were employed on the calculated values of the SSI. After the values for SSI were calculated, all of the values were scanned for unrealistically high values. Several exceedingly high values were discovered. In that case, a Skew T, Log P diagram was plotted manually following the method described in the previous section. All values obtained from the Skew T, Log P charts were close to the calculated values, and the values were accepted as plausible. To further ensure that the calculated SSI values were correct, a program called SHARP was used. In this program, the user inputs the 850-mb and 500-mb temperatures and dewpoints, and the corresponding Skew T, Log P diagram is drawn; among the calculations SHARP performs is the SSI. Although the SSI values were very similar, the values from SHARP were three-tenths higher than the calculated values, on average.

3.6 Research Approach

Upon receiving the necessary surface and upper air data from AFCCC, the data was sorted and manipulated into a more useful format. Through a series of FORTRAN-77 programs, the applicable months, hours, and pressure levels were extracted from the main data set. After sorting the data, performing several quality control checks, taking out missing data, and matching the days for which all variables were available, there were only 17 years of upper air data remaining. Of these 17 years, 15 were used to build the regression model and two were used in the validation. For every predictor included in the regression model, five to ten observations should be used in the validation, ten being ideal (Reynolds, 1997). Since the regression model included five predictors, fifty

observations were preferable for the validation. Two years of observations fulfilled this requirement.

The first variables that were calculated were the u and v components of the 850-mb and 500-mb winds. The wind data was given in terms of wind speed and direction, which were then converted into their respective u and v components using equations (12-15). As was mentioned in Chapter 2, the winds prior to 1956 were reported to the nearest integer multiple of 22.5 degrees. Depending on the direction and magnitude of the wind, large errors may occur. Therefore, as in Neumann's study, the years before 1956 were not included in this study. The regression was run with and without the years before 1956. However, using only the years from 1956 and later yielded better results.

The relative humidity is recorded as the radiosonde ascends through the atmosphere. Usually, the data is recorded in 50-mb increments. Sometimes, however, data at additional pressure levels is recorded. For the sake of simplicity, only the standard pressure levels (600 mb, 650 mb, 700 mb, 750 mb, and 800 mb) were used in the mean RH computation, which was computed by equation (17).

Computing the SSI was quite complicated since several layers and properties of the atmosphere had to be accounted for. As was mentioned in a previous section, the FORTRAN-77 program that was used in the computation of the SSI is included in Appendix D.

Finally, the frequency of thunderstorms was computed from 46 years of surface data using equation (12). Neumann characterized an afternoon thunderstorm as the event that either a thunderstorm or thunder heard was reported between the hours of 1000-2200

EST (Neumann, 1971: 2). Following his technique of using a binomial probability distribution for thunderstorm occurrence, if this criterion was met, a 1 was recorded for the day. But if no thunderstorm activity was reported, a 0 was recorded. Then, for each day of the convective season, all the 1s were tallied, and this total was used in equation (11) to compute the 15-day moving averages. Having calculated all the necessary parameters required by the algorithm, the data was split by month, thus yielding a data set for each month of the convective season. The data was then imported into Statistix[©], a powerful statistical software package. The combinations of variables on the right sides of equations (1-5) were easily calculated in Statistix[©]. Then, the initial linear regression was performed. In this initial regression, the transformed variables (as dictated by the second and third-order polynomials) were regressed against the binomial probability distribution as described previously. Thunderstorm frequency was regressed against day number in equation (5), following Neumann's work (Neumann, 1971: 9). As was mentioned previously, when using the REEP method, it is possible to obtain probabilities outside the range from 0 to 1 (Wilks, 1995: 183). After making the aforementioned variable transformations, approximately 5% of the polynomials, which are simply probabilities, fell outside this range. These polynomials were trimmed by assigning a 0 to negative polynomial values and a 1 to polynomial values greater than one.

From the tabular output from Statistix[©], the coefficients for each predictor were extracted and substituted as the constants in equations (1-5). Although the complete set of polynomials is given in Appendix E, the polynomials for May are shown below:

$$\begin{aligned}
F(X_1) = & 0.21097 + 0.01240S + 0.01375T + 0.0005471ST - 0.00006877S^2 \\
& + 0.0001359T^2 - 0.00002525S^3 - 0.00002006S^2T + 0.00003838ST^2 \\
& - 0.000008826T^3
\end{aligned} \tag{18}$$

$$\begin{aligned}
F(X_2) = & 0.14938 + 0.00659U + 0.01027V + 0.0001674UV + 0.0003401U^2 \\
& + 0.00006027V^2 - 0.000009582U^3 - 0.000007148U^2V - 0.000004493UV^2 \\
& - 0.000004883V^3
\end{aligned} \tag{19}$$

$$F(X_3) = -0.04712 - 0.00284RH + 0.0003155RH^2 - 0.000002521RH^3 \tag{20}$$

$$F(X_4) = 0.35954 - 0.06246SSI + 0.00247SSI^2 \tag{21}$$

$$F(X_5) = -0.98936 + 0.01267DAY - 0.00002717DAY^2 \tag{22}$$

Once the first regression had been done and the constants had been substituted into equations (1-5), the second regression was performed. In this regression, the five polynomials were regressed against the same binomial probability distribution. As was done in the first regression, the coefficients were taken from the output table and were substituted into equations (6-10), thereby yielding the five monthly probability equations. This set of probability, or predictor, equations was found to be:

$$P(\text{May}) = -0.19725 + 0.53729F(X_1) + 0.39294F(X_2) + 0.50858F(X_3) + 0.50194F(X_4) - 0.07365F(X_5) \quad (23)$$

$$P(\text{Jun}) = -0.72963 + 0.56485F(X_1) + 0.56162F(X_2) + 0.39101F(X_3) + 0.37465F(X_4) + 0.93634F(X_5) \quad (24)$$

$$P(\text{Jul}) = -1.10442 + 0.83962F(X_1) - 0.01663F(X_2) + 0.54726F(X_3) + 0.41546F(X_4) + 1.64595F(X_5) \quad (25)$$

$$P(\text{Aug}) = -0.69208 + 0.52504F(X_1) + 0.56613F(X_2) + 0.46575F(X_3) + 0.47203F(X_4) + 0.61684F(X_5) \quad (26)$$

$$P(\text{Sep}) = -0.54080 + 0.36607F(X_1) + 0.71709F(X_2) + 0.77712F(X_3) + 0.08617F(X_4) + 0.98095F(X_5) \quad (27)$$

In order to run the Neumann-Pfeffer Thunderstorm Index, the variables' constants in each month's polynomials and predictor equations must be assembled in a one-column format. Thus, for each month, there were 36 constants: 10 for the u component, 10 for the v component, 4 for RH, 3 for SSI, 3 for day number, and 6 for the final predictor equation. To clarify, the constants from equations (1-5) and equations (6-10), a total of 180 constants, were used as input for the NPTI code. Appendix C contains the list of constants for all five months for Neumann's study and for this study (Neumann, 1971: 39).

To compare the two algorithms, the NPTI was run using an independent sample of two years, 1983 and 1985. There were only four years from which to select the independent years: 1983, 1985, 1987, and 1988. It would not be appropriate to validate using the years that were used to build the regression model; this would have biased the results. Those four years were the only years not included by Neumann for which data was available to use in the validation. The years 1983 and 1985 were randomly chosen.

The two years were tested using Neumann's constants, the set of retuned constants, and persistence. After the NPTI had been run, thunderstorm probabilities for each day were reported as the output. Then, several statistics were computed using another FORTRAN-77 program. These statistics include the HR, FAR, POD, TS-Yes, TS-No, SS against persistence, and bias ratio. A description of these statistics is presented in Chapter 4.

4. Statistical Analysis

4.1 Introduction

Statistical analysis was perhaps the most important task involved in this project. It was the statistical analysis that gave meaning to the results obtained from the study. In this chapter, various statistics and their significances are examined.

4.2 The 2 X 2 Contingency Table

A 2 X 2 contingency table is a statistical tool used to show the number of occurrences and forecasts of a certain event. The table is composed of four quadrants. Appendix H contains an example of a 2 X 2 contingency table. The upper left quadrant (A) represents the event of a thunderstorm being observed given that one was also forecast. The upper right quadrant (B) represents the event of a thunderstorm not being observed given that a thunderstorm was forecast. The lower left quadrant (C) gives the event of a thunderstorm being observed given that a thunderstorm was not forecast. Finally, the lower right quadrant (D) is the event of a thunderstorm not being observed given that one was not forecast. For a completely accurate forecast method, entries of 0 would be shown in the lower left and upper right quadrants of the table. In other words, every time a thunderstorm was forecast, one was observed; and for every time a thunderstorm was not forecast, none was observed (Wilks, 1995: 238-239). Because no forecast method is perfect, statistical measures of accuracy are used to determine the values of different forecast methods.

4.3 Measures of Accuracy

Accuracy refers to the “average correspondence between individual forecasts and the events they predict” (Wilks, 1995: 236). Many measures of accuracy can be used to examine categorical “yes/no” forecasts. Some commonly used measures of accuracy are hit rate (HR), probability of detection (POD), false alarm rate (FAR), threat score (TS), and skill score (SS). The bias ratio (B) is also a useful measure; in the context of thunderstorm forecasting, the bias value reports whether thunderstorms are being over-forecast or under-forecast (Wilks, 1995: 239-241). Each of these measures is described below. It should be noted that in this study, all measures of accuracy are reported as percentages. To account for this, the formulas given, with the exception of the SS formula, should be multiplied by 100%.

4.3.1 Hit Rate (HR)

Hit rate, also known as proportion correct, is perhaps the most intuitive measure used to describe the accuracy of categorical forecasts. The hit rate is the fraction of the N forecasting occasions when the event was correctly forecast. The best possible hit rate is one, and the worst possible is zero. Thus, the hit rate percent ranges from 100% to 0%. The formula for computing the hit rate, which is obtained from the 2 X 2 contingency table, is shown below (Wilks, 1995: 240):

$$HR = \frac{A + D}{N} \quad (28)$$

Where:

A, D = values from the contingency table

N = total number of forecasting occasions (A+B+C+D)

4.3.2 False Alarm Rate (FAR)

The false alarm rate is the proportion of forecast events that fail to occur. The FAR is equivalent to the conditional probability of an event not being observed given that the event was forecast. Since the FAR has a negative connotation, smaller values are preferable. To that end, the best FAR is zero and the worst is one. A 2 X 2 contingency table can be used to compute the FAR using the formula (Wilks, 1995: 241):

$$FAR = \frac{B}{A + B} \quad (29)$$

4.3.3 Probability of Detection (POD)

The probability of detection is the ratio of correct forecasts of a certain event to the total times the event was observed. The POD is equivalent to the conditional probability of the event being forecast given that the event occurred. As is the case with the HR, the best possible value is one; the worst possible value is zero. Once again using the contingency table as an example, the formula is (Wilks, 1995: 240):

$$POD = \frac{A}{A + C} \quad (30)$$

4.3.4 Threat Score (TS)

When the event being forecast, the “yes” event, occurs less frequently than the “no” event, a commonly used measure is the threat score, also called the critical success index (CSI). The TS-Yes is the ratio of correct “yes” forecasts to the total number of occasions that the event was forecast or observed. In other words, the TS-Yes is the number of correct “yes” forecasts (Wilks, 1995: 240).

Using similar logic, a slightly different statistic was also calculated in this study, the TS-No. This represents the number of correct “no” forecasts. Both formulas are shown below:

$$TSYES = \frac{A}{A + B + C} \quad (31)$$

$$TSNO = \frac{D}{B + C + D} \quad (32)$$

4.3.5 Skill Score (SS)

Forecast skill is a term used to describe the relative accuracy of a set of forecasts with respect to some standard, or reference, forecasts. Wilks lists several possible sources of these reference forecasts including climatology, random forecasts, and

persistence. For this study, persistence was used as reference forecasts. Skill score is interpreted as the percentage improvement over the reference forecasts. While this measure of accuracy is not derived directly from the contingency table, the measure of accuracy A is taken from the table. It should be noted that this formula is initially calculated as a percentage; the other formulas were computed as ratios and multiplied to obtain percentages. The formula used to compute the SS is shown below (Wilks, 1995: 237):

$$SS = \frac{A - A(REF)}{A(PER) - A(REF)} \times 100\% \quad (33)$$

Where:

A = a particular measure of accuracy (HR, POD, FAR, TS-Yes, or TS-No)

A (ref) = the same measure of accuracy for the reference forecasts

A (per) = the same measure of accuracy for perfect forecasts

If A = A (per), the SS is 100%, the maximum value. If A = A (ref), the SS is 0%; in this case, the new forecasts are no better than the reference forecasts. If the new forecasts are not better than the reference forecasts, the SS is negative. And if the new forecasts are better than the reference forecasts, the SS is positive (Wilks, 1995: 237-238).

4.3.6 Bias Ratio (B)

The bias, which compares the average forecast to the average observation, is generally expressed as a ratio. The bias is the ratio of the number of “yes” forecasts to the number of “yes” observations. An event is said to be unbiased if the number of “yes” forecasts equals the number of “yes” observations, yielding a bias of one. A bias of one means that for every time a thunderstorm was observed, one had been forecast. A bias of less than one means that the event was forecast less often than it was observed; this is under-forecasting. Conversely, a bias of greater than one means that the event was forecast more often than it was observed; this is over-forecasting. Below is the formula used to calculate the bias ratio (Wilks, 1995: 241):

$$B = \frac{A + B}{A + C} \quad (34)$$

4.4 Pearson Correlation Coefficient

The Pearson Correlation Coefficient, R, is often used to describe the association between the predictors. R is bounded between -1 and $+1$. If R equals -1 , there is “perfect, negative linear association” between the predictors (Wilks, 1995: 46). Conversely, if the value of R is $+1$, there is “perfect, positive linear association” between the predictors (Wilks, 1995: 46). To give an example, consider a scatter plot. If the line of best fit is drawn through the data points on the scatter plot, the slope of the line can be either positive or negative. If the slope is positive, the R value is positive; if the slope is

negative, the R value is negative. Furthermore, the better the fit of the line, the closer the Pearson correlation coefficient is to 1 (either positive or negative). If the value of R is squared, the measure takes on a different meaning. The square of R specifies the amount of variance that can be explained by each predictor. Appendix G provides a plot of the Pearson correlation coefficients between the predictor functions and afternoon thunderstorm occurrence.

4.5 Test for Significance in the 2 X 2 Contingency Table

Above, several measures of accuracy and their importance were discussed. However, before these measures can be meaningful, the rows and columns of the 2 X 2 contingency table must be shown to be related or dependent. If the rows and columns do not exhibit dependence, then what appears to be a good relationship between thunderstorm forecasting and thunderstorms being observed could actually be due to chance (Kalbfleisch, 1979: 148). These random relationships are to be avoided. To that end, a test for independence must be performed to show that this dependent relationship exists among the rows and columns.

The Fisher-Irwin test, run in Statistix®, was used to show this dependent relationship between thunderstorm forecasts and thunderstorm observations. Before performing the test, the assumption was that the forecasts and observations were independent of each other. This hypothesis must be rejected in order to prove the dependence required for statistical significance. The test was run using a level of significance of five-percent (Sachs, 1984: 370-372). Under this level of significance, if

the computed p-value was less than 0.05, the assumption of independence was rejected in favor of dependence.

5. Results, Conclusions, and Recommendations

5.1 Introduction

This chapter includes a discussion of the forecast verification used in this study. Neumann's method of verification is briefly discussed. The measures of accuracy mentioned in Chapter 4 were computed for the current NPTI (with and without the subjective correction factor), the upgraded NPTI (with and without the subjective correction factor) and persistence forecasting. Five categories of cutoff percentages for thunderstorm forecasting were used: 35%, 40%, 45%, 50%, and 55%. The statistical measures of accuracy are discussed by cutoff percentages. Then the skill scores against persistence are discussed for the measures of accuracy for each NPTI. Finally, some recommendations for further research projects are mentioned.

5.2 Neumann's Forecast Verification

After using 13 years to build his model, Neumann used the dependent data set, the years 1957-1969, to verify the NPTI. As discussed in Chapter 2, he determined that forecast probabilities of greater than .050 were too low, and forecast probabilities of less than 0.50 were too high. He noted that forecast probabilities close to 0.50 were usually correct. While Neumann did not know the reason for this bias, he speculated that it was probably associated with the fitting of the polynomial equations (1-5) using a binomial probability distribution (Neumann, 1971: 30). He did point out, however, that June was the only month for which this bias was exhibited. Table 5 below summarizes the results of Neumann's dependent verification (Neumann, 1971: 30) of the NPTI.

Table 5. Verification of Current NPTI based on Dependent Data Set
for June (Neumann, 1971: 30)

Forecast probability	Number of thunderstorms forecast	Number of thunderstorms observed	Total number of cases	Observed occurrence rate
.00 to .05	1	64	65	.015
.06 to .15	2	20	22	.090
.16 to .25	2	23	25	.080
.26 to .35	5	29	34	.147
.36 to .45	21	45	66	.318
.46 to .55	26	27	53	.490
.56 to .65	34	18	52	.654
.66 to .75	35	8	43	.822
.76 to .85	17	3	20	.855
.86 to .95	3	1	4	.750
.00 to .95	146	238	384	.380

Neumann also tested the NPTI using a one-year independent data set, the year 1970. While he did not discuss any results of this verification, he stated that the results of the independent verification were similar to the results of the dependent verification. He stressed that more independent data must be used to fully assess the performance of the NPT (Neumann, 1971: 29-30).

5.3 Using 24-Hour Persistence for Forecasting a Thunderstorm

Day-to-day persistence, or 24-hour persistence, was considered in this study. Persistence forecasting means that if a thunderstorm is observed on one day, a thunderstorm is forecast for the next day. On the other hand, if no thunderstorm is

observed on a certain day, no thunderstorm is forecast for the next day. For the lower cutoff percentages for forecasting a thunderstorm, persistence forecast worse than either NPTI. However, at the 50% cutoff, persistence approaches the HR and TS-Yes scores for both NPTIs and exceeds the POD of either NPTI. The FAR of persistence is still higher than either version of the NPTI, and the TS-No values are similar; persistence produced unbiased forecasts. At the 50% and 55% cutoff levels, forecasting persistence produces more accurate forecasts than either NPTI. This will be addressed later in this chapter.

Table 6 presents the statistics when persistence was used as a forecast method in this study. This table and the tables that follow also list the inputs for building the 2 X 2 contingency tables.

Table 6. Using persistence as a tool for forecasting thunderstorms

Measure of accuracy	Value
HR	66%
TS-yes	46%
TS-no	50%
POD	63%
FAR	36%
Bias	0.99
A	87
B	50
C	52
D	108
N	297

5.4 Using 35% as Cutoff for Forecasting a Thunderstorm

A p-value of 0.00 was obtained from the Fisher-Irwin test, which was run in Statistix©. Therefore, the requirement for dependence in the 2 x 2 table was met, and the statistics in this discussion are meaningful.

The hit rate for the NPTI (with correction factor) was 74%, and the hit rate for the upgraded NPTI (with correction factor) was 73%. By including the subjective correction factor, the hit rates for both the current NPTI and the upgraded NPTI increased by less than 1%. The TS-Yes was clearly higher, with or without the correction factor, for the current NPTI. On the other hand, the TS-no was slightly better for the upgraded NPTI regardless of whether or not the correction factor was included. The current NPTI also had a large advantage over the upgraded NPTI for the POD. However, the upgraded NPTI was superior to the current NPTI for the FAR. The bias ratio of the current NPTI indicates over-forecasting. The upgraded NPTI was virtually unbiased. All things considered, using 35% as the cutoff for forecasting a thunderstorm, the current NPTI seems to have a slight, although virtually insignificant, advantage over the upgraded NPTI. Both NPTIs, though, are better than forecasting persistence, which is supported through the skill scores. As was discussed in Chapter 4, a positive SS represents improvement over the reference forecasts, persistence. Although the range of skill scores is from 6% to 65%, both NPTIs consistently out-perform persistence. Table 7 gives the statistics using 35% as the cutoff percentage for forecasting a thunderstorm. The SS using persistence as the reference forecasts are listed in all of the following tables.

Appendix I includes histograms for the HR, POD, FAR, TS-Yes, and TS-No at all five cutoff levels.

Table 7. Statistics for current NPTI and upgraded NPTI using 35% as cutoff

	NPTI with correction	NPTI without correction	Upgraded NPTI with correction	Upgraded NPTI without correction
HR	74%	73%	71%	71%
SS(HR)	24%	21%	21%	18%
TS-yes	63%	62%	55%	55%
SS(TS-yes)	31%	30%	20%	20%
TS-no	53%	53%	55%	55%
SS(TS-no)	6%	6%	14%	14%
POD	87%	87%	70%	70%
SS(POD)	65%	65%	24%	24%
FAR	31%	32%	29%	29%
SS(FAR)	8%	6%	14%	13%
Bias	1.27	1.28	0.99	0.99
A	130	130	105	105
B	59	60	42	43
C	19	19	44	44
D	89	88	106	105
N	297	297	297	297

5.5 Using 40% as Cutoff for Forecasting a Thunderstorm

For this category, too, a p-value of 0.00 was computed from the Fisher-Irwin test. Hence, the forthcoming statistics are valid.

The hit rate of the current NPTI, both with and without the correction factor, is slightly higher than the upgraded NPTI. Like in the 35% cutoff category, the TS-yes was better for the current NPTI than it was for the upgraded NPTI. But the TS-no for the upgraded NPTI edged the current NPTI by 2%. The current NPTI, with or without the correction factor, yielded a POD of 80%. This POD is 16% higher than the upgraded NPTI. The FAR of the upgraded NPTI, however, is better than that of the current NPTI. Interestingly, the current NPTI tended to over-forecast thunderstorms by the approximately the same margin that the upgraded NPTI tended to under-forecast thunderstorms at the 40% cutoff. Even though the upgraded NPTI produced more desirable TS-no and FAR statistics, the current NPTI was shown to be clearly better overall. Once again, forecasting persistence was the least desirable forecasting method, which is evident by the relatively high skill scores of both NPTIs. Table 8 outlines the statistics discussed here.

5.6 Using 45% as Cutoff for Forecasting a Thunderstorm

In this case, too, the computed p-value was found to be less than 0.05, the desired level of significance. Therefore, the results obtained from the statistical analysis are significant.

Once again, the current NPTI was shown to have a slightly higher HR than the upgraded NPTI. As has been the trend, the TS-Yes values for the current NPTI are better, while the upgraded NPTI has TS-No values that are higher than the current NPTI.

Table 8. Statistics for current NPTI and upgraded NPTI using 40% as cutoff

	NPTI with correction	NPTI without correction	Upgraded NPTI with correction	Upgraded NPTI without correction
HR	74%	73%	71%	71%
SS(HR)	24%	21%	18%	18%
TS-yes	60%	60%	53%	53%
SS(TS-yes)	26%	26%	13%	13%
TS-no	56%	56%	58%	58%
SS(TS-no)	12%	12%	16%	16%
POD	80%	80%	64%	64%
SS(POD)	46%	46%	3%	3%
FAR	29%	29%	25%	25%
SS(FAR)	11%	11%	19%	19%
Bias	1.12	1.13	0.86	0.86
A	119	119	96	96
B	48	49	32	32
C	30	30	53	53
D	100	99	116	116
N	297	297	297	297

Amazingly, the current NPTI had a POD of 16 percentage points better than the upgraded NPTI. But the upgraded NPTI had by far the lower FAR. Both the current NPTI and the upgraded NPTI tended to under-forecast thunderstorms at this cutoff percentage. As has been the case in the previous categories, the upgraded NPTI has an advantage in the TS-No values and FAR, and the current NPTI seems to produce better results than either persistence or the upgraded index. It is interesting, however, that persistence does perform better than the upgraded NPTI in the category of POD, as is indicated by the negative SS. Also, the SSs are getting lower, which means the gap between the NPTIs

and persistence is narrowing. Below, Table 9 lists the statistics for each NPTI at the 45% cutoff level.

Table 9. Statistics for current NPTI and upgraded NPTI using 45% as cutoff

	NPTI with correction	NPTI without correction	Upgraded NPTI with correction	Upgraded NPTI without correction
HR	72%	72%	71%	71%
SS(HR)	18%	18%	12%	12%
TS-yes	56%	56%	51%	51%
SS(TS-yes)	19%	19%	6%	6%
TS-no	57%	57%	59%	59%
SS(TS-no)	14%	14%	16%	16%
POD	70%	70%	60%	60%
SS(POD)	19%	19%	-14%	-14%
FAR	27%	27%	23%	23%
SS(FAR)	14%	14%	19%	19%
Bias	0.96	0.97	0.77	0.77
A	105	105	89	89
B	38	39	26	26
C	44	44	60	60
D	110	109	122	122
N	297	297	297	297

5.7 Using 50% as Cutoff for Forecasting a Thunderstorm

The Fisher-Irwin test produced a value of 0.00, which is representative of dependence among the rows and columns in the 2 X 2 contingency table. This means that the statistics calculated are valid.

As is shown in Table 10, the current NPTI and the upgraded NPTI yielded basically the same HR. The difference of 1% is negligible. Even though the current NPTI still had a slight edge over the upgraded NPTI in terms of TS-Yes, the difference continues to decrease. The upgraded NPTI has a TS-No value of 59%, and that beat the current NPTI. The current NPTI again had a better POD than the upgraded NPTI, but persistence had the best POD by eight percentage points. The results for the FAR were still lower for the upgraded NPTI. However, persistence out-performs either NPTI, with or without the correction factor, by a large margin in the POD category; the skill scores are still decreasing. Thus, persistence forecasting is performing almost as accurately as either NPTI. Finally, both NPTIs under-forecast thunderstorms. For the 50% cutoff category, the only advantages the current NPTI has over the upgraded NPTI is TS-Yes and POD, and these advantages are slight. Overall, the two indices are comparable at this cutoff percentage, and persistence performs about as well as either NPTI.

5.8 Using 55% as Cutoff for Forecasting a Thunderstorm

As a result of the p-values computed in the Fisher-Irwin test being less than 0.05, the results in this section should be accepted as meaningful.

For both the current NPTI and the upgraded NPTI, the HRs were approximately the same. The TS-Yes values for the current index were also slightly higher than that of the upgraded NPTI; the upgraded NPTI had a slightly higher TS-No percent, but it was

Table 10. Statistics for current NPTI and upgraded NPTI using 50% as cutoff

	NPTI with correction	NPTI without correction	Upgraded NPTI with correction	Upgraded NPTI without correction
HR	70%	70%	69%	69%
SS(HR)	12%	12%	9%	9%
TS-yes	49%	49%	45%	45%
SS(TS-yes)	6%	6%	-2%	-2%
TS-no	58%	57%	59%	59%
SS(TS-no)	16%	14%	16%	16%
POD	58%	58%	50%	50%
SS(POD)	-14%	-14%	-30%	-30%
FAR	24%	24%	18%	18%
SS(FAR)	19%	19%	24%	24%
Bias	0.77	0.77	0.62	0.62
A	87	87	75	75
B	27	28	17	17
C	62	62	74	74
D	121	120	131	131
N	297	297	297	297

not significantly different from that of the current NPTI. The PODs and the FARs of the two indices were also comparable. The current NPTI and the upgraded NPTI were both guilty of under-forecasting. In this cutoff category, the upgraded NPTI seems to perform approximately as accurately as the current NPTI, but persistence forecasting performed better than either version of the NPTI. This is evident by the negative skill scores for TS-Yes and POD. In the other categories in which the SS did not become negative, the values were still decreasing. Table 11 highlights these statistics.

Table 11. Statistics for current NPTI and upgraded NPTI using 55% as cutoff

	NPTI with correction	NPTI without correction	Upgraded NPTI with correction	Upgraded NPTI without correction
HR	66%	66%	64%	64%
SS(HR)	0%	0%	0%	0%
TS-yes	38%	38%	35%	35%
SS(TS-yes)	-15%	-15%	-15%	-15%
TS-no	57%	57%	56%	56%
SS(TS-no)	14%	14%	16%	16%
POD	42%	42%	38%	38%
SS(POD)	-57%	-57%	-59%	-59%
FAR	18%	19%	20%	20%
SS(FAR)	28%	27%	31%	31%
Bias	0.51	0.52	0.48	0.48
A	62	62	57	57
B	14	15	14	14
C	87	87	92	92
D	134	133	134	134
N	297	297	297	297

5.9 Conclusions

While the upgraded NPTI consistently proved to be better in categories such as TS-No and FAR, the current NPTI performed better in HR, TS-Yes, and POD. As the percentage cutoff category for forecasting a thunderstorm increased, the upgraded NPTI seemed to produce slightly better statistics. However, so did persistence. In fact, the SS progressively decreased for every statistic as the percentage cutoff level increased. At the 50% cutoff level, persistence produced results very similar to the current NPTI, and at the 55% level, persistence performed better than the current index. At both the 50% level

and the 55% level, the skill scores for TS-Yes and POD went negative, which supports the notion that persistence out-performed either NPTI at those cutoff levels. The upgraded NPTI tended to under-forecast at all five cutoff percentages. The crossover from under-forecasting to over-forecasting in the current NPTI occurred at the 45% cutoff, while persistence, as mentioned earlier, was virtually unbiased.

5.10 Recommendations

For operational use it is recommended that the current NPTI, for lack of significant improvement, continue to be used for forecasting thunderstorms at Cape Canaveral, Florida. Because persistence performed as well as, or in some cases out-performed, both the current NPTI and the upgraded NPTI at higher cutoff percentages, a more accurate forecasting method must be developed and implemented immediately.

5.11 Suggestions for Future Research

As this project was unfolding, several other potential research ideas came to mind. The K Index, rather than the SSI, should be considered as an input variable into the NPTI. In a previous study, the K Index was the only stability index to have “modest utility in discriminating convective activity in the vicinity of KSC (Kennedy Space Center)” (Bauman et al., 1996). Another idea might be to use the wind speed and direction, as reported in knots and degrees, as input variables instead of the u and v components. Depending on the direction and magnitude of the wind, some error could potentially occur in the conversions to u and v components. If speed and direction were used as predictors rather than converting to u and v components, it is quite possible to trim some of this potential error.

Another consideration should be the implementation of a different type of regression known as logistics regression. The advantage to using logistics regression is that the probabilities yielded from the polynomial equations (1-5) are bounded between 0 and 1 (Wilks, 1995: 183). In the REEP approach, used by Neumann and in this study, these probabilities are not guaranteed to be bounded, and that must be taken into account. For example, in Neumann's study, he bounded his probabilities with a series of ellipses, as described in Chapter 2. In this study, any probabilities outside the range from 0 to 1 were trimmed, as discussed in Chapter 3. The use of logistics regression would eliminate the problem of bounding the probabilities. The possibilities are endless.

Appendix A. Wind rose plots for all months

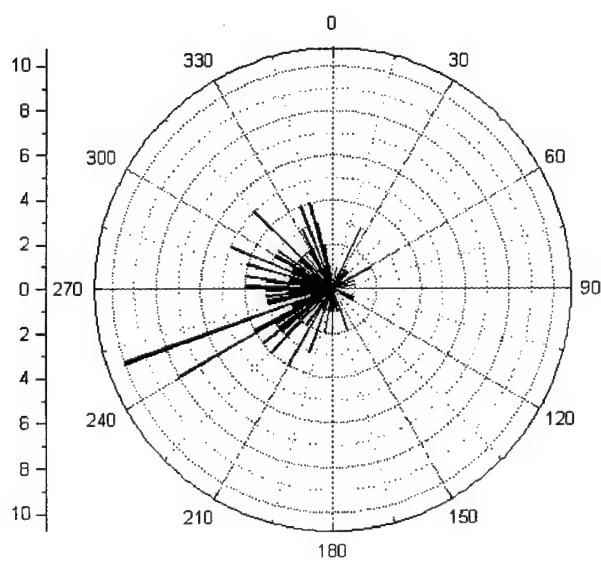
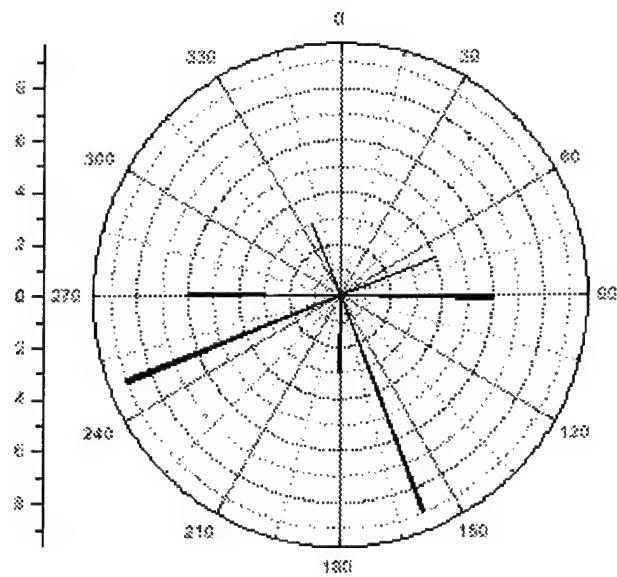


Figure A-1. 500-mb wind roses for May for the years a.) before 1956 and b.) from 1956 and later

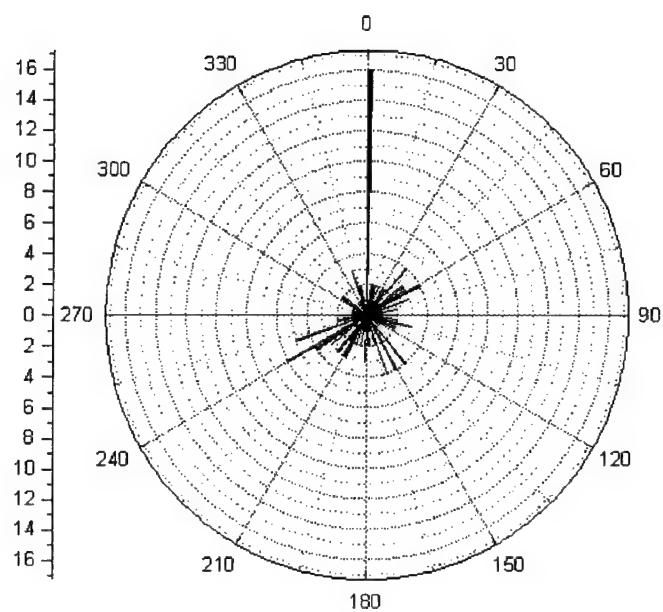
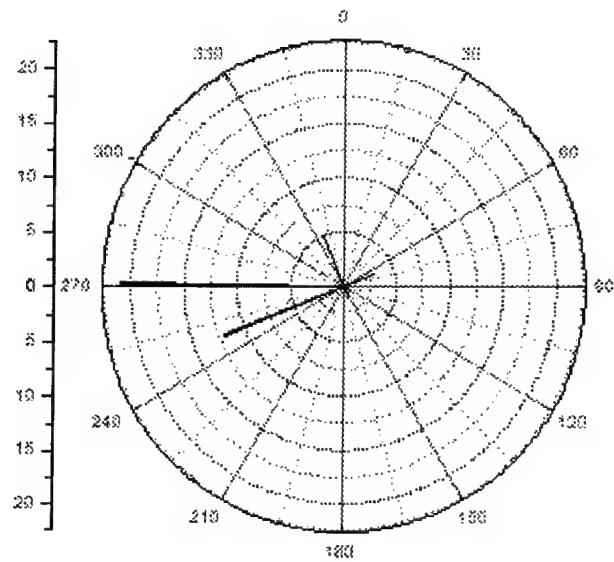


Figure A-2. 850-mb wind rose plots for May for the years a.) before 1956 and b.) from 1956 and later

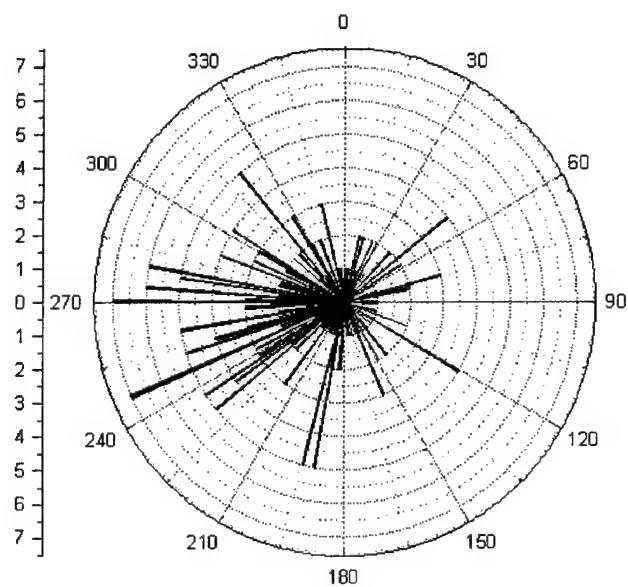
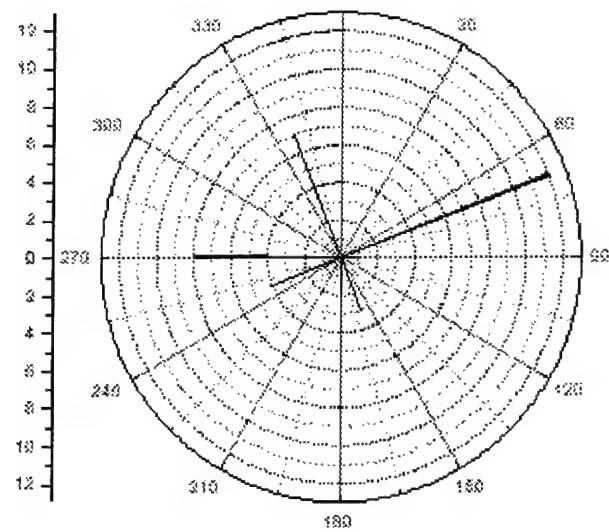


Figure A-3. 500-mb wind rose plots for June for the years a.) before 1956 and b.) from 1956 and later

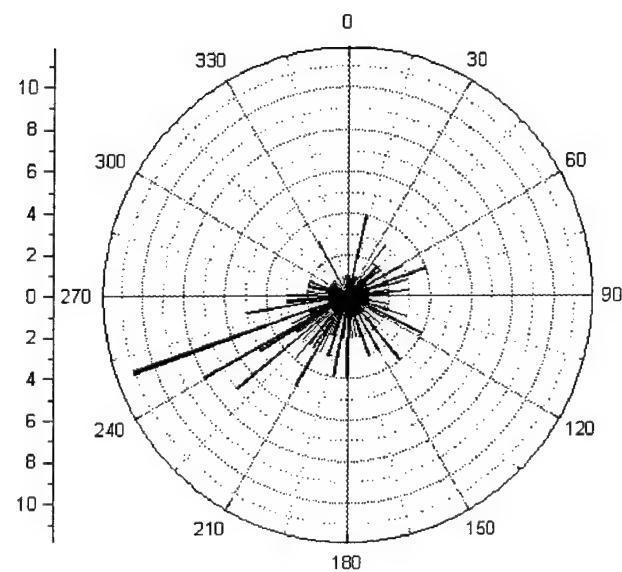
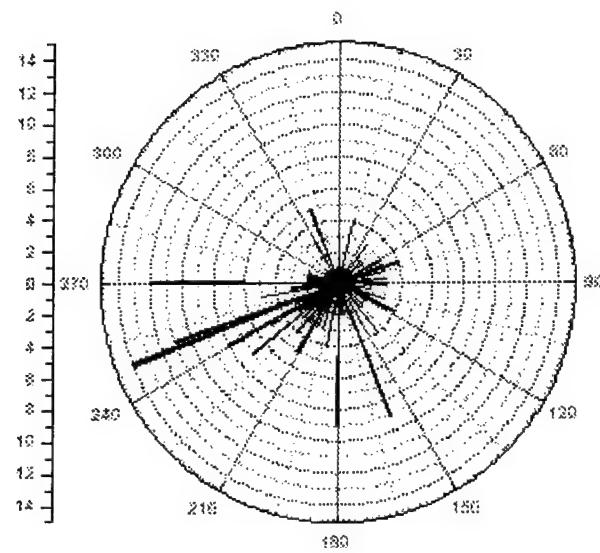


Figure A-4. 850-mb wind rose plots for June for the years a.) before 1956 and b.) from 1956 and later

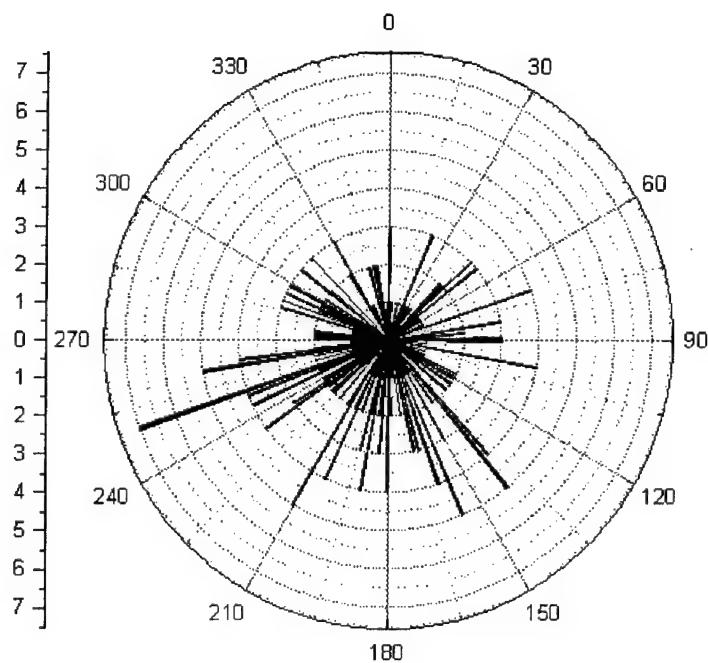
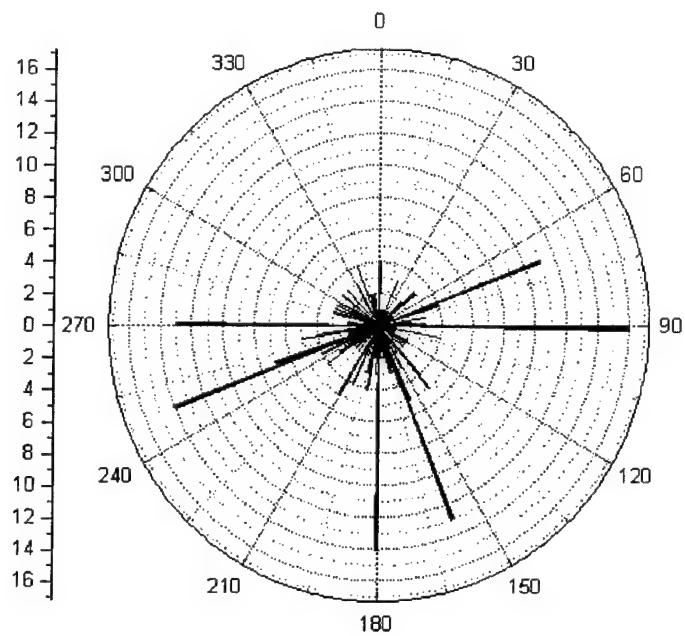


Figure A-5. 500-mb wind rose plots for July for the years a.) before 1956 and b.) from 1956 and later

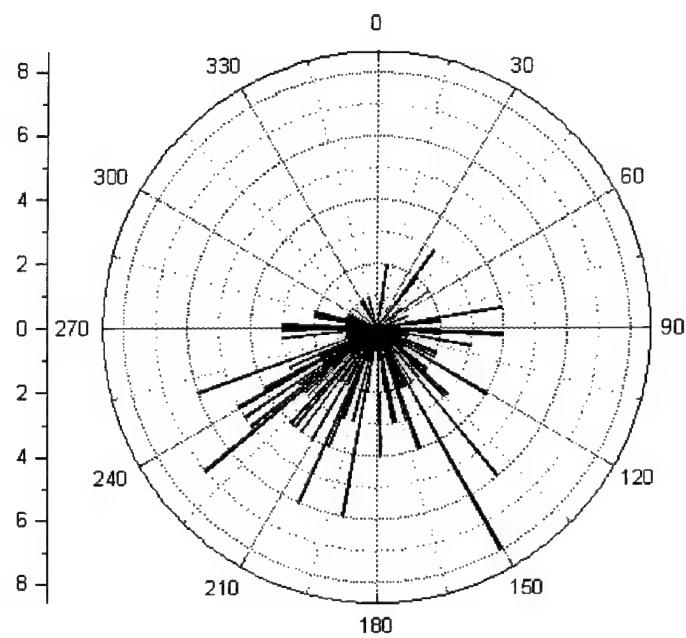
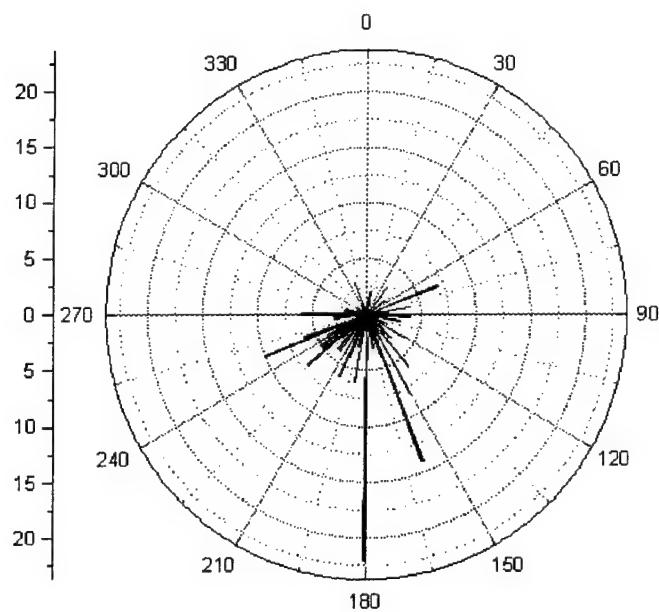


Figure A-6. 850-mb wind rose plots for July for the years a.) before 1956 and b.) from 1956 and later

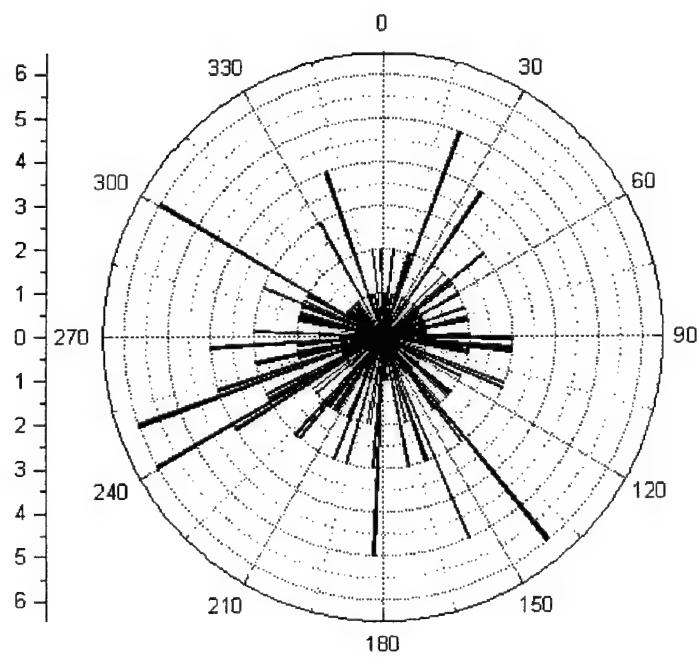
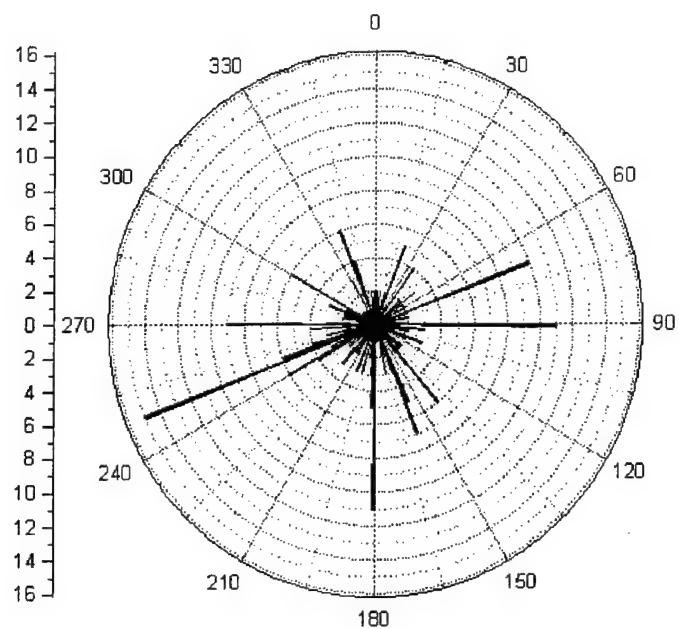


Figure A-7. 500-mb wind rose plots for August for the years a.) before 1956 and b.) from 1956 and later

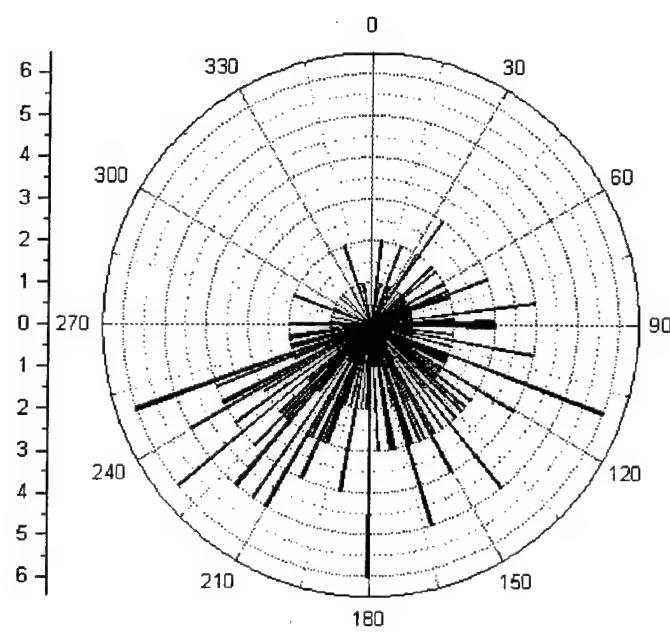
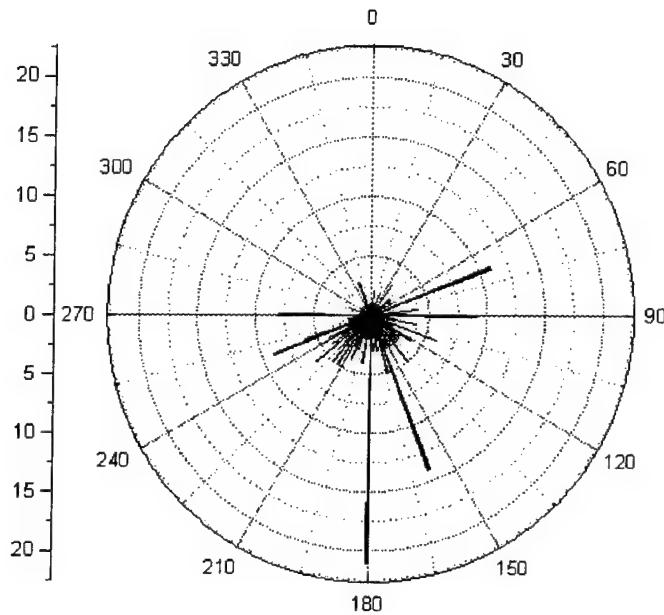


Figure A-8. 850-mb wind rose plots for August for the years a.) before 1956 and b.) from 1956 and later

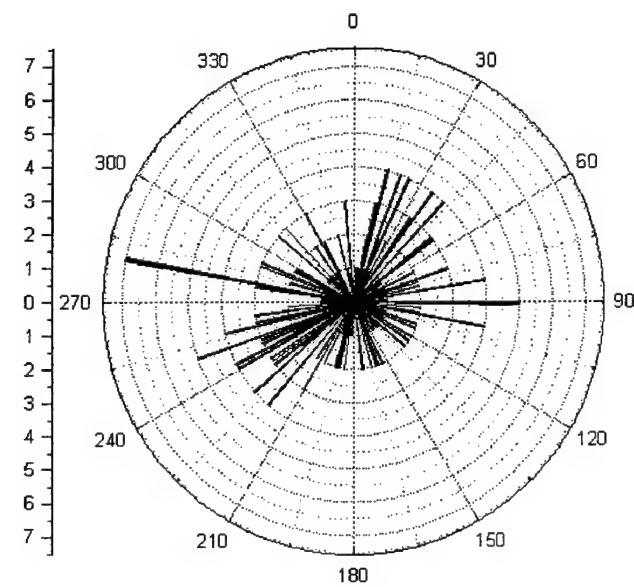
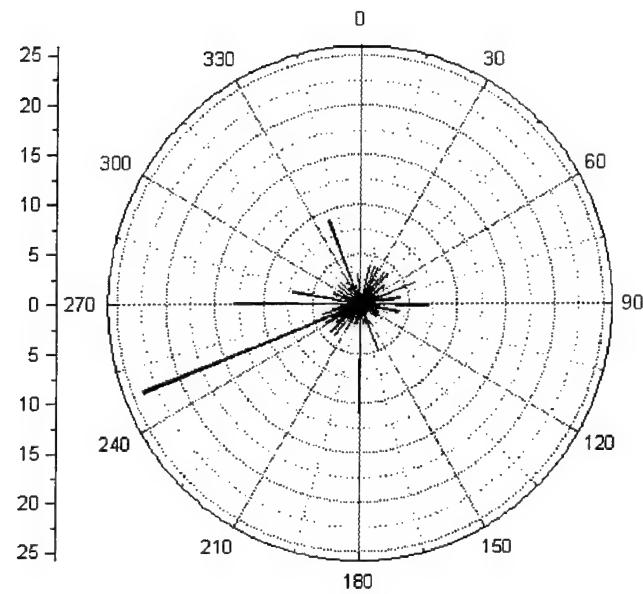


Figure A-9. 500-mb wind rose plots for September for the years a.) before 1956 and b.) from 1956 and later

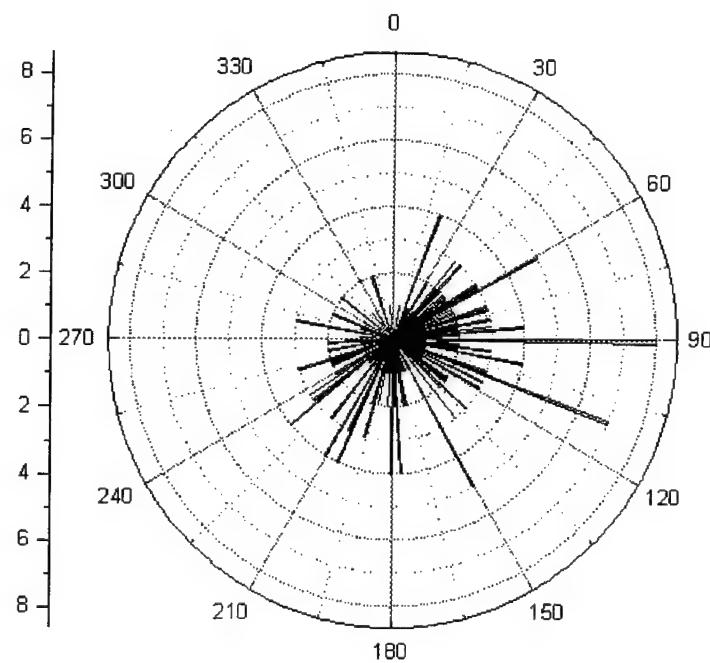
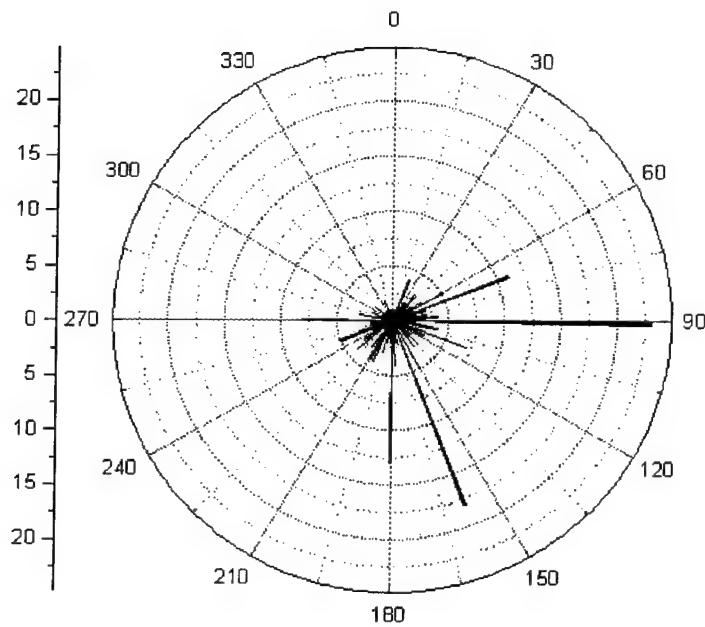
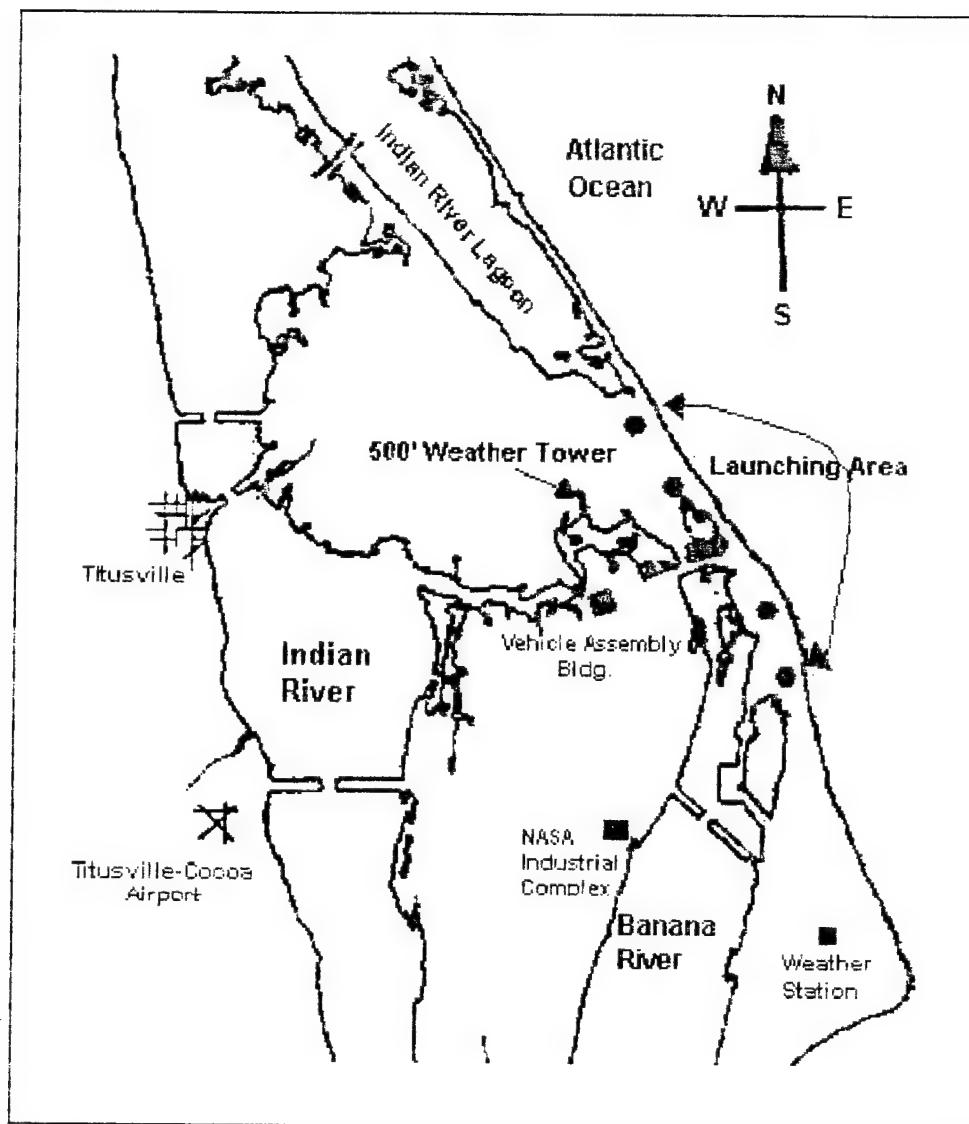


Figure A-10. 850-mb wind rose plots for September for the years a.) before 1956 and b.) from 1956 and later

Appendix B. Station IDs and geographic changes of Cape Canaveral, Florida

Inclusive Dates	ID	Geographic Location
June 1950-16 March 1978	KXMR	Weather Station A (Cape Canaveral Air Station)
17 March 1978-31 July 1980	KX68	Shuttle Landing Facility on Cape Kennedy Space Center
1 August 1980-10 February 1993	KX68	Weather Station B (still on KSC)
11 February 1993-19 May 1993	KQCH	Weather Station B
20 May 1993-16 June 1993	KKSC	Weather Station B
17 June 1993-Present	KTTS	Weather Station B

Map of Cape Canaveral area



Appendix C. Input Constants for current NPTI (Neumann, 1971: 39)

F(X1) May	F(X1) June	F(X1) July	F(X1) Aug	F(X1) Sept
+0.1787416E+0	+0.3326784E+0	+0.4307867E+0	+0.3627524E+0	+0.2816768E+0
+0.1074020E-1	+0.2172438E-1	+0.4366697E-1	-0.3272211E-1	+0.1256518E-1
+0.1365651E-1	+0.2162950E-1	+0.1055475E-1	+0.1085207E-1	+0.5804331E-2
+0.4523660E-3	+0.3762057E-3	-0.3983282E-5	-0.5623188E-4	+0.1096534E-3
-0.1802959E-3	-0.6835820E-3	-0.3116466E-3	+0.1038914E-2	0.2671097E-3
+0.3397793E-3	+0.2579027E-3	-0.1888946E-2	-0.3726892E-3	+0.1469291E-4
-0.1051838E-4	+0.1179004E-5	-0.5616631E-4	-0.3354727E-4	-0.1099520E-4
-0.3954366E-4	+0.1437934E-5	+0.7757704E-4	-0.1055251E-3	+0.2925611E-5
+0.3376410E-4	-0.3373770E-4	-0.5417381E-4	-0.6772392E-5	+0.3228711E-5
+0.1677435E-5	-0.2199710E-4	+0.3519052E-4	+0.1606764E-4	-0.3225703E-5
F(X2) May	F(X2) June	F(X2) July	F(X2) Aug	F(X2) Sept
+0.1206249E+0	+0.2927882E+0	+0.4145883E+0	+0.3932798E+0	+0.2527479E+0
+0.1080646E-1	+0.2638450E-1	+0.3166340E-1	+0.3119719E-1	+0.1084204E-1
+0.1001964E-1	+0.1023307E-1	-0.7151265E-3	+0.2545731E-2	+0.3136786E-2
+0.2794513E-3	+0.3206674E-3	+0.5390950E-3	+0.1592548E-3	+0.1899334E-3
-0.1012098E-3	+0.7055071E-4	+0.4251009E-4	+0.9662810E-4	-0.2175208E-3
+0.1964561E-3	+0.1576005E-3	-0.5091109E-4	+0.2887853E-4	-0.3547892E-4
-0.1929388E-5	-0.3090318E-4	-0.2425546E-4	-0.3745136E-4	-0.5449895E-5
-0.1095389E-4	-0.1422489E-4	+0.1581160E-4	-0.1717338E-4	-0.4427336E-5
-0.6512555E-5	+0.5588606E-5	-0.2172134E-4	-0.1704165E-4	+0.6122512E-5
-0.1931907E-5	-0.9225416E-5	-0.1060904E-4	+0.4082921E-5	+0.5412232E-5
F(X3) May	F(X3) June	F(X3) July	F(X3) Aug	F(X3) Sept
+0.1037449E+0	+0.1350110E+0	-0.1029031E+0	+0.2562494E+1	+0.1736004E+0
-0.1196854E-1	-0.1999291E-1	-0.2906759E-2	-0.1702073E+0	-0.1918291E-1
+0.4832994E-3	+0.8150660E-3	+0.4229306E-3	+0.3551389E-2	+0.6220713E-3
-0.3570444E-5	-0.6342578E-5	-0.3308301E-5	-0.2161341E-4	-0.4414412E-5
F(X4) May	F(X4) June	F(X4) July	F(X4) Aug	F(X4) Sept
+0.4273235E+0	+0.6102192E+0	+0.6177575E+0	+0.5271789E+0	+0.4078606E+0
-0.7480216E-1	-0.8066767E-1	-0.6421018E-1	-0.3530199E-1	-0.6376678E-1
+0.3056711E-2	+0.2403726E-2	+0.1310411E-2	-0.1094883E-2	+0.2571961E-2
F(X5) May	F(X5) June	F(X5) July	F(X5) Aug	F(X5) Sept
-0.5430778E+0	-0.1323037E+0	+0.9355280E+0	-0.4163536E+0	+0.3758034E+1
+0.6855607E-2	+0.1070858E-2	-0.3771816E-2	+0.1394724E-1	-0.2287890E-1
-0.1053707E-4	+0.2308962E-4	+0.6918595E-5	-0.4493190E-4	+0.3598785E-4
Poly(May)	Poly(June)	Poly(July)	Poly(Aug)	Poly(Sept)
-0.1589528E+0	-0.5556250E+0	-0.5553775E+0	-0.4622971E+0	-0.6182956E+0
+0.5503053E+0	+0.6102450E+0	+0.6370509E+0	+0.6391629E+0	+0.5269289E+0
+0.3738171E+0	+0.4851770E+0	+0.4154169E+0	+0.4061392E+0	+0.6065540E+0
+0.3233246E+0	+0.3646010E+0	+0.4982033E+0	+0.4244231E+0	+0.5538999E+0
+0.5656907E+0	+0.354164E+0	+0.4217904E+0	+0.5676596E+0	+0.4831459E+0
+0.2053246E-1	+0.6391500E+0	+0.2361394E+0	+0.6062162E-1	+0.1294910E+1

Note: When inputting the coefficients above into the NPTI FORTRAN code, they should be entered in the following manner: F(X1), F(X2), F(X3), F(X4), F(X5), and Poly(month) for May, then for June, July, Aug, and Sept.

Input Constants for the upgraded NPTI

F(X1) May	F(X1) June	F(X1) July	F(X1) Aug	F(X1) Sept
+0.2109700E+0	+0.3451200E+0	+0.4749800E+0	+0.3408300E+0	+0.2605200E+0
+0.1240000E-1	+0.2334000E-1	+0.3617000E-1	+0.2742000E-1	+0.1002000E-1
+0.1375000E-1	+0.1770000E-1	+0.6880000E-2	+0.5330000E-2	+0.6350000E-2
+0.5471000E-3	-0.1184000E-3	+0.7066000E-3	+0.6436000E-3	-0.2240000E-4
-0.6872000E-4	-0.1696000E-3	-0.3756000E-3	+0.9636000E-3	-0.4581000E-4
+0.1359000E-3	-0.8486000E-4	-0.1750000E-2	-0.5583000E-4	-0.1197000E-3
-0.2525000E-4	-0.8390000E-5	-0.5156000E-4	-0.1548000E-4	-0.4608000E-5
-0.2006000E-4	-0.1500000E-4	+0.5850000E-4	-0.5628000E-4	-0.5226000E-6
+0.3838000E-4	-0.2890000E-4	-0.7541000E-4	-0.9240000E-4	+0.6208000E-6
-0.8826000E-5	+0.1006000E-5	+0.4666000E-4	+0.2258000E-4	-0.1190000E-5
F(X2) May	F(X2) June	F(X2) July	F(X2) Aug	F(X2) Sept
+0.1493800E+0	+0.2944100E+0	+0.4373300E+0	+0.3475000E+0	+0.2307500E+0
+0.6590000E-2	+0.2202000E-1	+0.2841000E-1	+0.2922000E-1	+0.7770000E-2
+0.1027000E-1	+0.8690000E-2	-0.3170000E-2	+0.8300000E-2	+0.7580000E-2
+0.1674000E-3	+0.7495000E-5	-0.5031000E-3	+0.3991000E-3	+0.2441000E-3
+0.3401000E-3	-0.1532000E-4	-0.1211000E-3	+0.1601000E-4	-0.1193000E-3
+0.6027000E-4	+0.6004000E-3	-0.2723000E-4	+0.9205000E-3	-0.3970000E-4
-0.9582000E-5	-0.2797000E-4	-0.3450000E-4	-0.4662000E-4	-0.5613000E-5
-0.7148000E-5	+0.7935000E-5	+0.6560000E-4	-0.2108000E-4	+0.4479000E-5
-0.4493000E-5	+0.9777000E-5	+0.7424000E-5	+0.2640000E-5	-0.6166000E-5
-0.4883000E-5	-0.9421000E-5	+0.3438000E-3	-0.4865000E-4	-0.4981000E-5
F(X3) May	F(X3) June	F(X3) July	F(X3) Aug	F(X3) Sept
-0.4712000E-1	+0.4668000E-1	-0.2065700E+0	+0.1147710E+1	+0.2912100E+0
-0.2840000E-2	-0.1220000E-1	+0.5740000E-2	-0.7456000E-1	-0.2537000E-1
+0.3155000E-3	+0.5782000E-3	+0.2609000E-3	+0.1670000E-2	+0.7215000E-3
-0.2521000E-5	-0.4265000E-5	-0.2505000E-5	-0.1041000E-4	-0.4963000E-5
F(X4) May	F(X4) June	F(X4) July	F(X4) Aug	F(X4) Sept
+0.3595400E+0	+0.5064000E+0	+0.5611800E+0	+0.4540500E+0	+0.3022900E+0
-0.6246000E-1	-0.6370000E-1	-0.6981000E-1	-0.5656000E-1	-0.3795000E-1
+0.2470000E-2	+0.1300000E-2	+0.7700000E-3	+0.2500000E-2	+0.7814000E-3
F(X5) May	F(X5) June	F(X5) July	F(X5) Aug	F(X5) Sept
-0.9893600E+0	-0.1459820E+1	+0.2209960E+1	-0.1882150E+1	-0.8897440E+1
+0.1267000E-1	+0.1609000E-1	-0.1722000E-1	+0.2404000E-1	+0.7792000E-1
-0.2717000E-4	-0.2989000E-4	+0.4182000E-4	-0.6073000E-4	-0.1634000E-3
Poly(May)	Poly(June)	Poly(July)	Poly(Aug)	Poly(Sept)
-0.4411000E-1	-0.6973400E+0	-0.1870750E+1	-0.6868300E+0	-0.7975000E+0
+0.3212000E+0	+0.5788900E+0	+0.8870000E+0	+0.5999000E+0	-0.3140000E+0
+0.2665000E+0	+0.5498100E+0	+0.1605000E-1	+0.5597700E+0	+0.7493400E+0
+0.51868900E+0	+0.4265900E+0	+0.5497100E+0	+0.3531700E+0	+0.7712000E+0
+0.6152100E+0	+0.4359100E+0	+0.4009200E+0	+0.5234900E+0	-0.1378000E-1
-0.5732900E+0	+0.7698300E+0	+0.3327770E+1	+0.6246000E+0	+0.1880990E+1

Note: When inputing the coefficients above into the NPTI FORTRAN code, they should be entered in the following manner: F(X1), F(X2), F(X3), F(X4), F(X5), and Poly(month) for May, then for June, July, Aug, and Sept.

- * This program was written by Christian S. Wohlwend, 2Lt, United States Air Force.
- * It was adapted for use on the data set used in this study.

PROGRAM SHOWALTER STABILITY INDEX

INTEGER HR, DAY, YR, MON, T8, TD8, T5, N, TD5

DOUBLE PRECISION T850	TEMPERATURE AT 850 MB
DOUBLE PRECISION TD850	DEWPOINT AT 850 MB
DOUBLE PRECISION T500	TEMPERATURE AT 500 MB
DOUBLE PRECISION TD500	DEWPOINT AT 500 MB
DOUBLE PRECISION TLCL	TEMPERATURE AT LFC
DOUBLE PRECISION PLCL	PRESSURE AT LCL
DOUBLE PRECISION E	VAPOR PRESSURE AT LCL
DOUBLE PRECISION EP	SATURATION VAPOR PRESSURE
DOUBLE PRECISION L	LATENT HEAT OF WATER VAPOR
DOUBLE PRECISION LP	LATENT HEAT OF PARCEL
DOUBLE PRECISION WLCL	MIXING RATIO AT 850 MB
DOUBLE PRECISION WP	SATURATION MIXING RATIO
DOUBLE PRECISION CP	SPECIFIC HEAT OF DRY AIR
DOUBLE PRECISION THETA_D	PARTIAL POTENTIAL TEMPERATURE
DOUBLE PRECISION THETA_SE	PSEUDO-EQUIVALENT POTENTIAL TEM
DOUBLE PRECISION THETAP	THETA D OF PARCEL
DOUBLE PRECISION THETA_EP	THETA SE OF PARCEL
DOUBLE PRECISION SSI	SHOWALTER STABILITY INDEX
DOUBLE PRECISION C	KELVIN CONVERSION
DOUBLE PRECISION K	RD/CP
DOUBLE PRECISION ERR	ERROR FUNCTION
DOUBLE PRECISION ERR_P	SECOND ERROR FUNCTION
DOUBLE PRECISION TP	TEMPERATURE GUESS
DOUBLE PRECISION TP2	SECOND TEMPERATURE GUESS
DOUBLE PRECISION DELTA_T	FRACTION OF TEMPERATURE GUESS
DOUBLE PRECISION EPSILON	ALLOWABLE ERROR
DOUBLE PRECISION ZERO	NUMBER ZERO

- * Define constants

CP=0.24

C=273.16

K=0.2854

EPSILON=0.05

ZERO=0.0

OPEN (UNIT=10,FILE='sepready',STATUS='OLD')

OPEN (UNIT=20,FILE='sepssi',STATUS='UNKNOWN')

- * Read in the data file

- * DO 3 N=1,1209

```

* DO 4 I=1,KE
* READ (10,*) HR,DAY,MON,YR,PRESS,T,TD
* WRITE (20,25) HR,DAY,MON,YR,PRESS,T,TD
*4 CONTINUE
*3 CONTINUE

DO 6 N=1,761

READ (10,26,END=999) HR,DAY,MON,YR,T8,TD8,T5,TD5

26 FORMAT (I2,1X,I2,1X,A3,1X,I4,2X,I3,1X,I2,1X,I3,1X,I3)

* Get rid of temps or dewpoints with 99 or 999 entries
* the strings 99 and 999 represent missing values

IF (T8 .NE. 99 .AND. TD8 .NE. 99 .AND. T5 .NE. 99) THEN
IF (T8 .NE. 999 .AND. TD8 .NE. 999 .AND. T5 .NE. 999) THEN

* Find the variables at the LCL

TLCL=(TD8-((0.212+0.001571*TD8-0.000436*T8)
$*(T8-TD8))+C)

T850=FLOAT(T8)+C

TD850=FLOAT(TD8)+C

T500=FLOAT(T5)+C

PLCL = 850.0*((TLCL/T850)**(1.0/K))

IF (TLCL.GE.C) THEN

E=(10.0**23.832241 - (5.02808*DLOG10(TLCL))
$-(1.3816*(10.0**(-7))*
$((10.0**11.334 - (0.0303998*TLCL))))*
$+(8.1328*(10.0**(-3))*
$((10.0**3.49149-(1302.8844/TLCL))))*
$-(2949.076/TLCL)))*

L=(597.3 - (0.564*(TLCL-C)))*

ELSE

E=(10.0**3.56654*DLOG10(TLCL)-
$(0.0032098*TLCL)-(2484.956/TLCL)
$+2.0702294)*

L=(597.3 - (0.574*(TLCL-C)))*

END IF

```

```

END IF

WLCL=((0.62197*E)/(PLCL-E))

THETA_D=(TLCL*((850.0/(PLCL-E))**K))

THETA_SE=THETA_D*(DEXP((L*WLCL)/(CP*TLCL)))

* Find TP500

TP = (C - 5.0)

DELTA_T = 0.05

EP = (10.0**((3.56654*DLOG10(TP)) - (0.0032098*TP))  

$- (2484.956/TP) + 2.0702294))  

LP = (597.3 - (0.574*(TP - C)))  

WP = ((0.62197*EP)/(500.0-EP))  

THETAP = (TP*((850.0/(500.0 - EP))**K))  

THETA_EP = THETAP*(DEXP((LP*WP)/(CP*TP)))  

ERR = (THETA_EP - THETA_SE)  

IF (ABS(ERR).LT.EPSILON) THEN  

TP500 = TP  

ELSE  

TP2 = TP + DELTA_T  

EP = (10.0**((3.56654*DLOG10(TP2)) - (0.0032098*TP2))  

$- (2484.956/TP2) + 2.0702294))  

LP = (597.3 - (0.574*(TP2 - C)))  

WP = ((0.62197*EP)/(500.0-EP))  

THETAP = (TP2*((850.0/(500.0 - EP))**K))  

THETA_EP = THETAP*(DEXP((LP*WP)/(CP*TP2)))  

ERR_P = (THETA_EP - THETA_SE)  

* WRITE(20,25) THETAP,THETA_EP,ERR_P,ERR  

IF (ABS(ERR_P).LT.EPSILON) THEN  

TP500= TP2

```

```
ELSE
IF ((ERR.LT.ZERO.AND.ERR_P.GT.ZERO).OR.
$(ERR.GT.ZERO.AND.ERR_P.LT.ZERO)) THEN
  DELTA_T = (0.5*(DELTA_T))
  GOTO 12
ELSE
  IF (ABS(ERR_P).LT.ABS(ERR)) THEN
    TP = TP2
    ERR = ERR_P
    GOTO 12
  ELSE
    DELTA_T = (-1.0*(DELTA_T))
    GOTO 12
  END IF
  END IF
  END IF
  END IF

* Calculate the SSI
SSI = (T500 - TP500)
SSI = (INT(SSI*100.0 +0.5))/100.0
WRITE (20.25) HR, DAY, MON, YR, SSI
END IF

25 FORMAT (I2,2X,I2,2X,A3,2X,I4,2X,F25.20)

9  CONTINUE
8  CONTINUE
6  CONTINUE
999 STOP

END
```

Appendix E. Polynomials

MAY

$$f(X_1) = 0.21097 + 0.01240S + 0.01375T + 0.0005471ST - 0.00006872S^2 + 0.0001359T^2 - 0.00002525S^3 - 0.00002006S^2T + 0.00003838ST^2 - 0.000008826T^3$$

$$f(X_2) = 0.14938 + 0.00659U + 0.01027V + 0.0001674UV + 0.0003401U^2 + 0.00006027V^2 - 0.000009582U^3 - 0.000007148U^2V - 0.000004493UV^2 - 0.000004883V^3$$

$$f(X_3) = -0.04712 - 0.00284RH + 0.0003155RH^2 - 0.000002521RH^3$$

$$f(X_4) = 0.35954 - 0.06246SSI + 0.00247SSI^2$$

$$f(X_5) = -0.98936 + 0.01267DAY - 0.00002717DAY^2$$

JUNE

$$F(X_1) = 0.34512 + 0.02334S + 0.01770T - 0.0001184ST - 0.0001696S^2 - 0.0008486T^2 - 0.000008390S^3 - 0.00001500S^2T - 0.00002890ST^2 + 0.000001006T^3$$

$$F(X_2) = 0.29441 + 0.02202U + 0.00869V + 0.000007495UV - 0.00001562U^2 + 0.0006004V^2 - 0.00002797U^3 + 0.000007935U^2V + 0.000009777UV^2 - 0.000009421V^3$$

$$F(X_3) = -0.04668 - 0.01220RH + 0.0005782RH^2 - 0.000004265RH^3$$

$$F(X_4) = 0.50640 - 0.06370SSI + 0.00130SSI^2$$

$$F(X_5) = -1.45982 + 0.01609DAY - 0.00002989DAY^2$$

JULY

$$F(X_1) = 0.47488 + 0.03617S + 0.00688T + 0.0007066ST - 0.0003756S^2 - 0.00175T^2 - 0.00005156S^3 + 0.00005850S^2T - 0.00007541ST^2 + 0.00004666T^3$$

$$F(X_2) = 0.43733 + 0.02841U - 0.00317V - 0.0005031UV - 0.0001211U^2 - 0.00002723V^2 - 0.0000034500U^3 + 0.00006560U^2V + 0.00007424UV^2 + 0.0003438V^3$$

$$F(X_3) = -0.20657 + 0.00574RH + 0.0002609RH^2 - 0.000002505RH^3$$

$$F(X_4) = 0.56118 - 0.06981SSI + 0.0007700SSI^2$$

$$F(X_5) = 2.20996 - 0.01722DAY + 0.00004182DAY^2$$

AUGUST

$$f(X_1) = 0.34083 + 0.02472S + 0.00533T + 0.0006436ST + 0.0009639S^2 - 0.00005583T^2 - 0.00001548S^3 - 0.00005628S^2T - 0.00009240ST^2 + 0.00002258T^3$$

$$f(X_2) = 0.34750 + 0.02972U + 0.00830V + 0.0003991UV + 0.00001601U^2 + 0.0009205V^2 - 0.00004662U^3 - 0.00002108U^2V + 0.000002640UV^2 - 0.00004865V^3$$

$$f(X_3) = 1.14771 - 0.07456RH + 0.00167RH^2 - 0.00001041RH^3$$

$$f(X_4) = 0.45405 - 0.05656SSI + 0.00250SSI^2$$

$$f(X_5) = -1.88215 + 0.02404DAY - 0.00006073DAY^2$$

SEPTEMBER

$$f(X_1) = 0.34083 + 0.02472S + 0.00533T + 0.0006436ST + 0.0009639S^2 - 0.00005583T^2 - 0.00001548S^3 - 0.00005628S^2T - 0.00009240ST^2 + 0.00002258T^3$$

$$f(X_2) = 0.34750 + 0.02972U + 0.00830V + 0.0003991UV + 0.00001601U^2 + 0.0009205V^2 - 0.00004662U^3 - 0.00002108U^2V + 0.000002640UV^2 - 0.00004865V^3$$

$$f(X_3) = 1.14771 - 0.07456RH + 0.00167RH^2 - 0.00001041RH^3$$

$$f(X_4) = 0.45405 - 0.05656SSI + 0.00250SSI^2$$

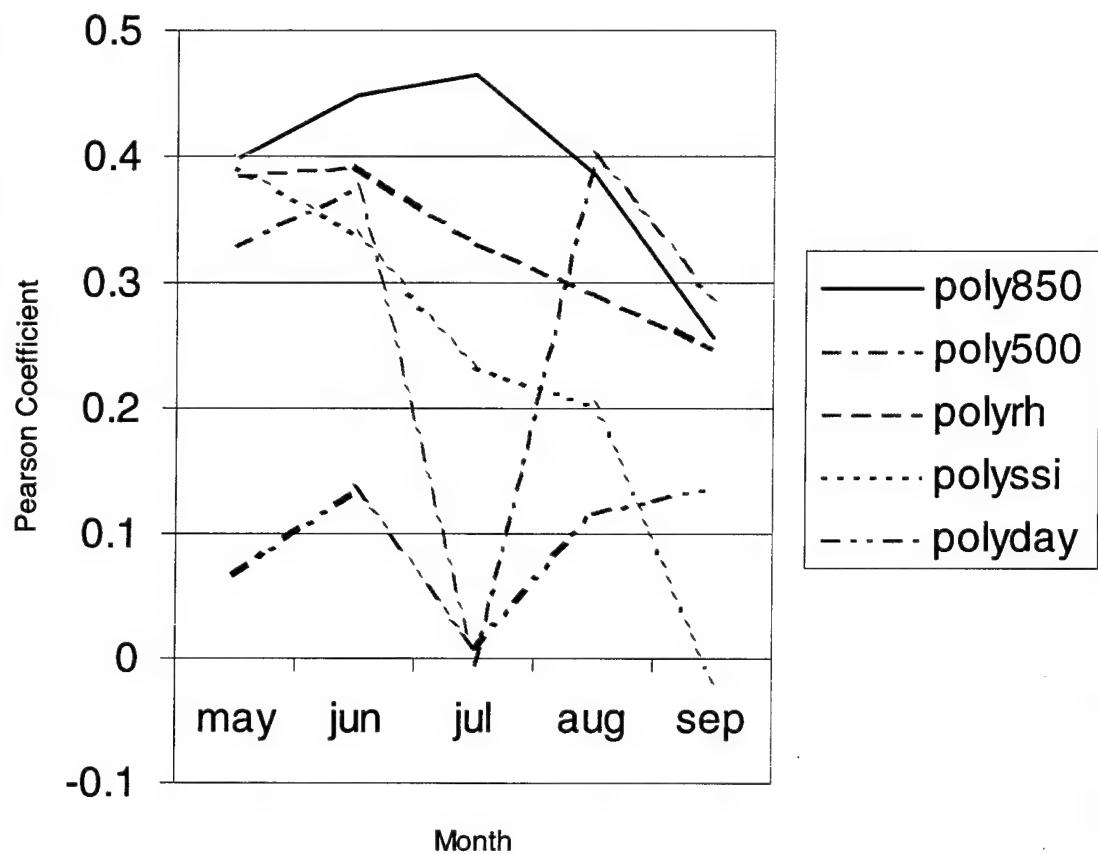
$$f(X_5) = -1.88215 + 0.02404DAY - 0.00006073DAY^2$$

Appendix F. P-values and chi-squared values for current NPTI and upgraded NPTI (with and without correction factor) at each cutoff percentage and for persistence

	P-value	Chi-squared value
35% Cutoff		
Current NPTI (with correction)	0.00	72.04
Current NPTI (without correction)	0.00	70.28
Upgraded NPTI (with correction)	0.00	52.62
Upgraded NPTI (without correction)	0.00	50.94
40% Cutoff		
Current NPTI (with correction)	0.00	67.88
Current NPTI (without correction)	0.00	66.07
Upgraded NPTI (with correction)	0.00	55.48
Upgraded NPTI (without correction)	0.00	55.48
45% Cutoff		
Current NPTI (with correction)	0.00	59.67

Current NPTI (without correction)	0.00	57.86
Upgraded NPTI (with correction)	0.00	55.63
Upgraded NPTI (without correction)	0.00	55.63
50% Cutoff		
Current NPTI (with correction)	0.00	50.60
Current NPTI (without correction)	0.00	48.75
Upgraded NPTI (with correction)	0.00	52.41
Upgraded NPTI (without correction)	0.00	52.41
55% Cutoff		
Current NPTI (with correction)	0.00	40.31
Current NPTI (without correction)	0.00	38.30
Upgraded NPTI (with correction)	0.00	33.84
Upgraded NPTI (without correction)	0.00	33.84
Persistence	0.00	28.49

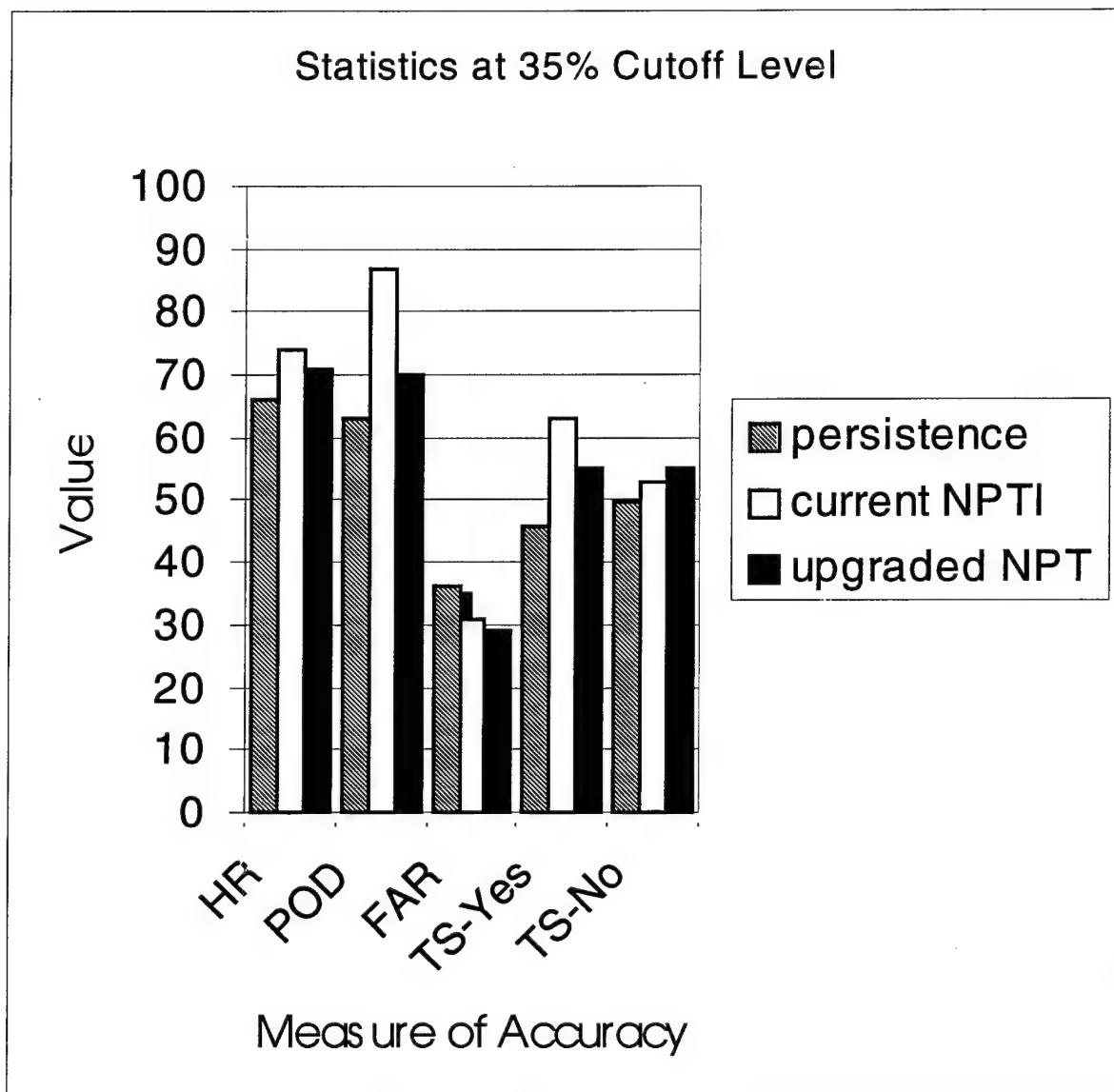
Appendix G. Pearson Correlation Coefficients
Between Predictor Functions and Afternoon
Thunderstorm Occurrence



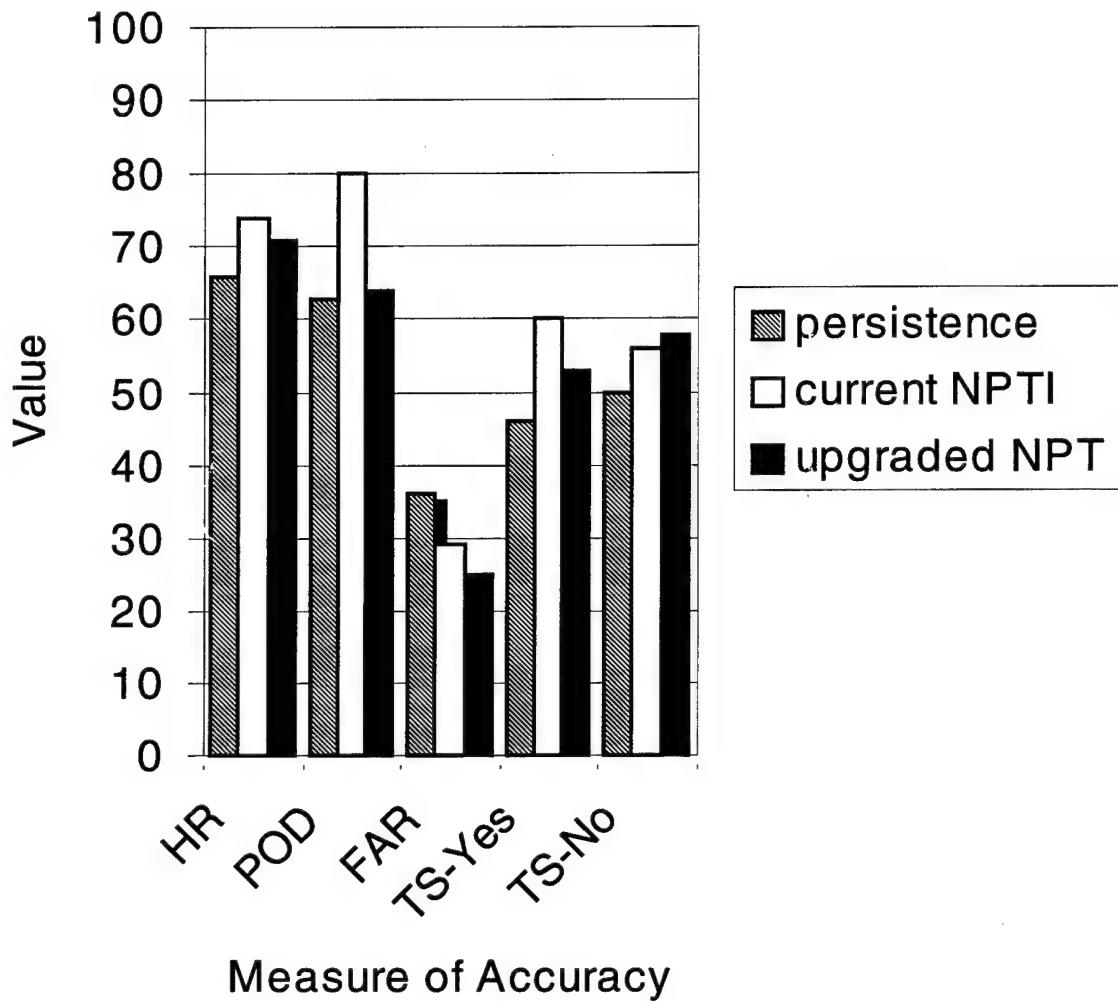
Appendix H. Example of 2 X 2 contingency table

		Thunderstorms Observed	
		YES	NO
Thunderstorm Forecast	YES	A	B
	NO	C	D

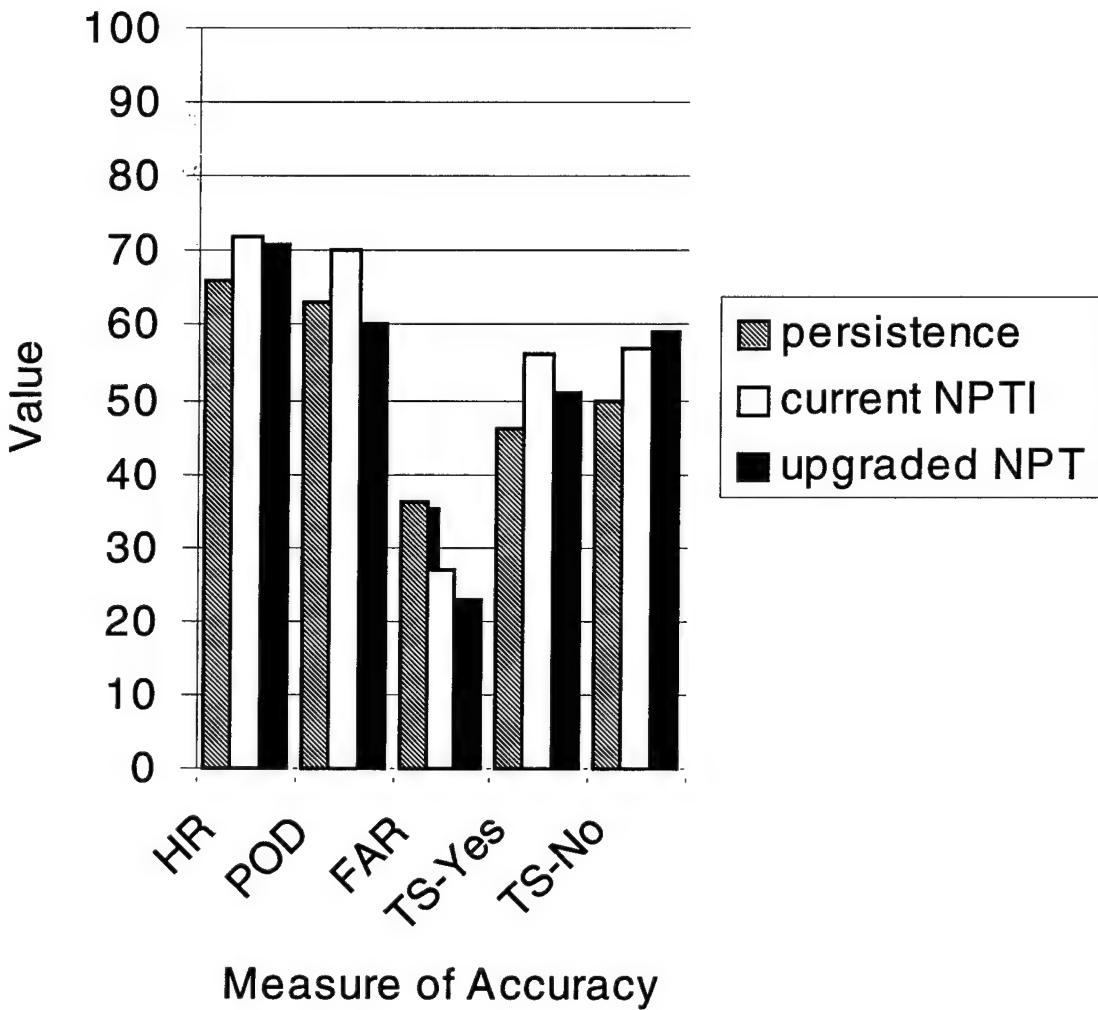
Appendix I. Statistics at all cutoff levels



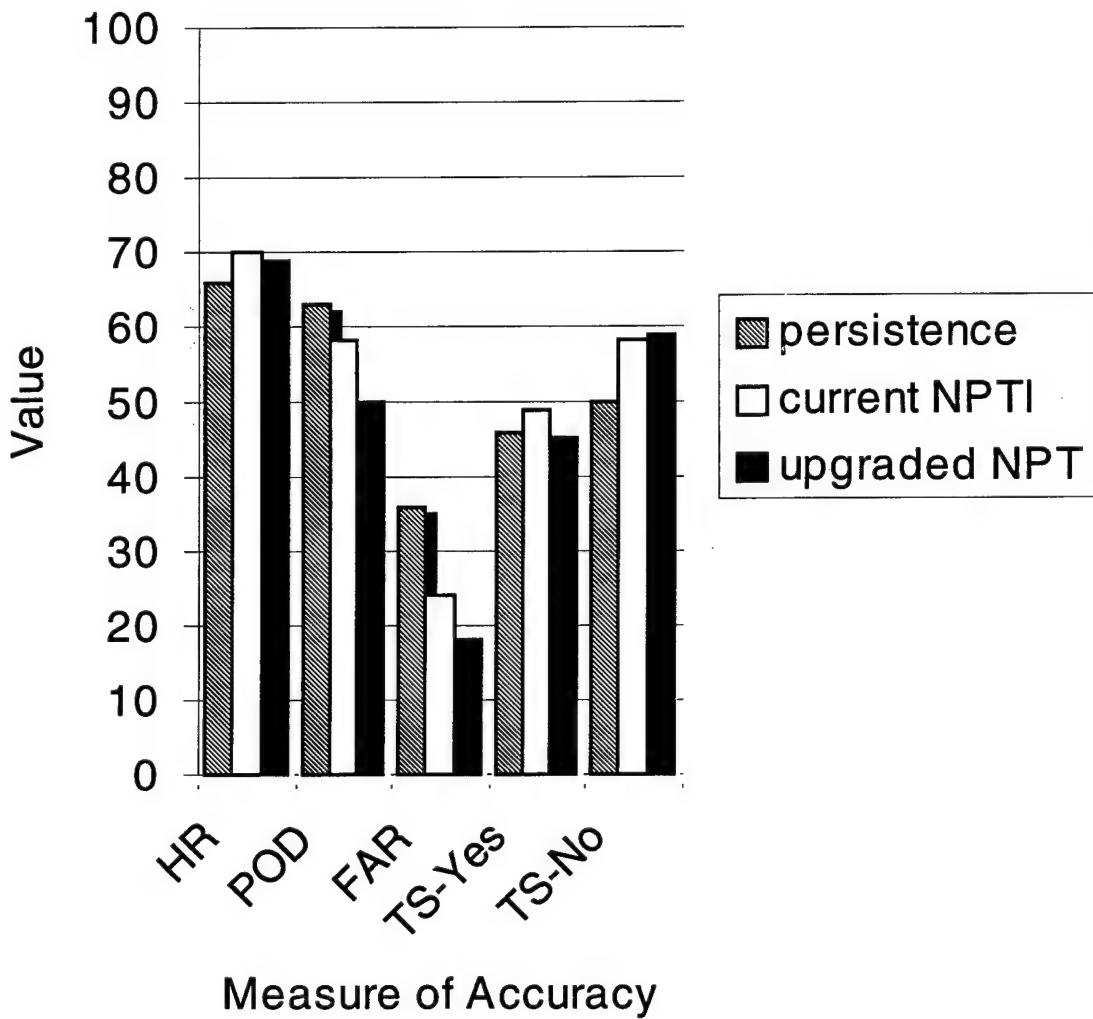
Statistics at 40% Cutoff Level



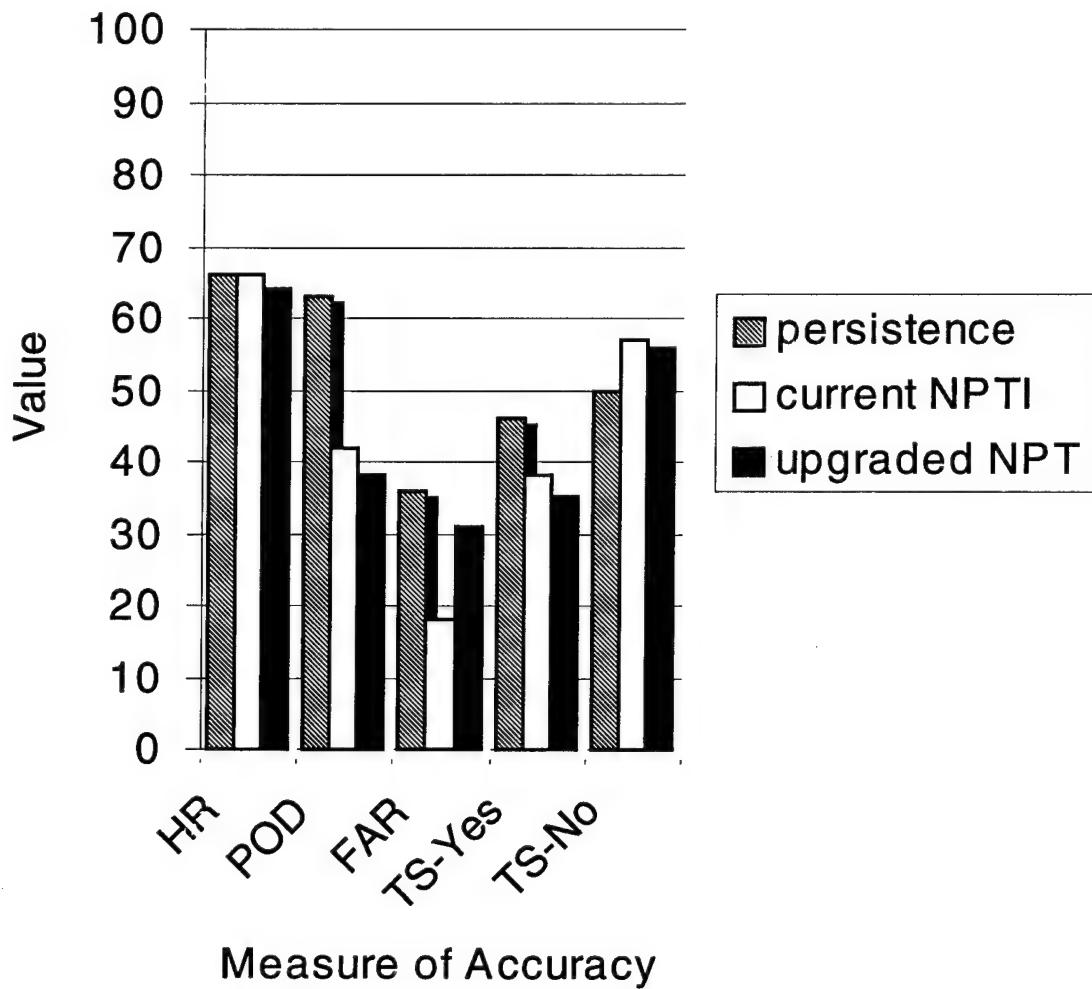
Statistics at 45% Cutoff Level



Statistics at 50% Cutoff Level



Statistics at 55% Cutoff Level



BIBLIOGRAPHY

Air Weather Service Technical Report AWS/TR-79/006: The use of the Skew T, Log P diagram in analysis and forecasting. Air Weather Service Technical Report-79/006.

Atkins, Nolan T. and Roger M. Wakimoto, 1997: Information of the synoptic-scale flow on sea breeze observations during CaPE. *Mon. Wea. Rev.*, **125**, 2112-2130.

Bauman, William H. III and Steven Businger, 1996: Nowcasting for space shuttle landings at Kennedy Space Center, Florida. *Bull. Amer. Meteor. Soc.*, 2295-2305.

Bauman, William H. III and Michael L. Kaplan, 1997: Nowcasting convective activity for space shuttle landings during easterly regimes. *Wea. Forecasting*, **12**, 78-107.

Cetola, Jeffrey D., 1997: A climatology of the sea breeze at Cape Canaveral, Florida. Unpublished Report. The Florida State University College of Arts and Sciences.

Clouse, Phillip G. and T. Johnathon Whiteside, 1996: Directory of Climatic Data Bases. Air Force Combat Climatology Center Technical Report-96/001.

Duffield, George F. and Gregory D. Nastrom, 1983: Equations and algorithms for meteorological applications in Air Weather Service. Air Weather Service Technical Report-83/001.

Frank, N.L., P.L. Moore, G.E. Fischer, 1967: Summer shower distribution of the Florida peninsula as deduced from digitized radar data. *J. Appl. Meteor.*, **6**, 309-316.

Golden, J. H., R. Serafin, V. Lilly, and J. Facundo, 1986: Atmospheric sounding systems. *Mesoscale Meteorology and Forecasting*, P. S. Ray, Ed., Amer. Metero. Soc., 50-70.

Hatcher, Larry and Edward J. Stepanski, 1994: A Step-by-step approach to using the SAS system for univariate and multivariate statistics: SAS Institute, Incorporated, 624 pp.

Hazen, D. S., W. P Roeder, B. F. Boyd, J. B. Lorens, and T. L. Wilde, 1995: Weather impact on launch operations at the Eastern Range and Kennedy Space Center. Sixth Conference on Aviation Weather Systems, Dallas, TX, Amer. Meteor. Soc., 270-275.

Jessup, Edward A., 1972: Interpretations of chaff trajectories near a severe thunderstorm. *Mon. Wea. Rev.*, **100**, 653-661.

Manobianco, J., J. W. Zack, and G. E. Taylor, 1996: Workstation-based real-time mesoscale modeling designed for weather support operations at the Kennedy Space Center and Cape canaveral Air Station. *Bull. Amer. Meteor. Soc.*, **77**, 653-672.

McGinley, John, 1986: Nowcasting mesoscale phenomena. *Mesoscale Meteorology and Forecasting*, P. S. Ray, Ed., Amer. Meteor. Soc., 657-688.

Neumann, Charles J., 1968: Frequency and duration of thunderstorms at Cape Kennedy, Part I. Weather Bureau Technical Memorandum SOS-2.

Neumann, Charles J., 1970: Frequency and duration of thunderstorms at Cape Kennedy, Part II. Weather Bureau Technical Memorandum SOS-6.

Neumann, Charles J., 1971: Thunderstorm forecasting at Cape Kennedy, Florida, utilizing multiple regression techniques. NOAA Technical Memorandum NWS SOS-8.

Rabayda, Allen. Captain, United States Air Force at Air Force Combat Climatology Center, Asheville, NC. Personal Correspondence. 7 January 1998.

Roeder, William. Chief of Operations Support Flight and Science And Technical Training Officer for 45th Weather Squadron, Patrick Air Force Base, FL. Personal interview. 18 June 1997.

Sachs, Lothar, 1984: *Applied Statistics: A Handbook of Techniques*. New York: Springer-Verlag.

Wallace, John M. and Peter V. Hobbs, 1977: *Atmospheric Science: An Introductory Survey*, Academic Press.

Weiss, Steven J., 1992: Some aspects of forecasting severe storms during cool-season return-flow episodes. *J. Appl. Meteor.*, **31**, 964-982.

Wilks, Daniel S, 1995: *Statistical Methods in the Atmospheric Sciences*, Academic Press.

Wilson, James W. and Daniel L. Megenhardt, 1997: Thunderstorm initiation, organization, and lifetime associated with Florida boundary layer convective lines. *Mon. Wea. Rev.*, **125**, 1507-1525.

Wohlwend, Christian, 1998: Improving Cape Canaveral's day two thunderstorm forecasting using meso-eta numerical model output. Unpublished report. The Air Force Institute of Technology.

Zhong, Shiyuan and Eugene S. Takle, 1992: An observational study of sea- and land-breeze circulations in an area of complex coastal heating. *J. Appl. Meteor.*, **31**, 1426-1438.

Zhong, Shiyuan and Eugene S. Takle, 1993: The effects of large-scale winds on the sea-land breeze circulations in an area of complex coastal heating. *J. Appl. Meteor.*, **32**, 1181-1195.

Vita

Lt Cindy L. Howell was born on 18 July 1974 in Enterprise, Alabama. She graduated as Valedictorian from Coffee Springs High School in 1992 and began undergraduate studies at Troy State University in Troy, Alabama. She graduated Cum Laude with a Bachelor of Arts degree in Mathematics in May 1995. She received her commission on 29 May 1996 upon completion of the Reserved Officers Training Corps (ROTC).

The assignment to the Air Force Institute of Technology was her first assignment. Her next assignment will be with the 355 Operations Support Squadron as a Weather Officer at Davis-Monthan Air Force Base in Tucson, Arizona.

Permanent Address: Route 1, Box 293
Coffee Springs, AL 36318

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)			2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
			March 1998	Master's Thesis
4. TITLE AND SUBTITLE			5. FUNDING NUMBERS	
NOW-CASTING THUNDERSTORMS AT CAPE CANAVERAL, FLORIDA, USING AN IMPROVED NEUMANN-PFEFFER THUNDERSTORM INDEX				
6. AUTHOR(S)				
Cindy L. Howell, 2Lt, USAF				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)			8. PERFORMING ORGANIZATION REPORT NUMBER	
Air Force Institute of Technology Advisor: Lt Col Mike Walters (DSN 785-3636 X 4681, mwalters@afit.af.mil) 2750 P Street Wright-Patterson AFB, OH 45433-7765			AFIT/GM/ENP/98M-05	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)			10. SPONSORING/MONITORING AGENCY REPORT NUMBER	
Mr. William Roeder (comm 407-494-7426, william-roeder@pafb.af.mil) 45th Weather Squadron 1201 Minuteman Street Patrick AFB, FL 32925-3238				
11. SUPPLEMENTARY NOTES				
12a. DISTRIBUTION AVAILABILITY STATEMENT			12b. DISTRIBUTION CODE	
Approved for public release; distribution unlimited				
13. ABSTRACT (Maximum 200 words)				
The Neumann-Pfeffer Thunderstorm Index (NPTI) is used daily by the 45th Weather Squadron during the convective season to estimate the probability of afternoon thunderstorms. The current NPTI, developed by Charles J. Neumann in the 1960s, is based on only 13 years of data taken over 30 years ago. The index was in desperate need of an upgrade. Following the multiple regression techniques outlined by Neumann, this thesis examines whether or not including additional data would improve the performance of the NPTI. After performing the multiple regressions and retuning the regression coefficients, both NPTIs were validated using a 2-year independent data set. Then, several measures of accuracy were computed to compare the current NPTI, the upgraded NPTI, and 24-hour persistence. At lower cutoff percentages, the current NPTI and the upgraded version performed quite similarly; persistence was the worst of the three methods. However, at higher cutoff percentages, persistence out-performed both versions of the NPTI. Both NPTIs still performed equally well. It is recommended that the current NPTI should continue to be used operationally since the upgraded NPTI did not offer any significant improvement. Furthermore, because persistence out-performed either NPTI at higher cutoff levels, a new forecast method must be developed and implemented immediately.				
14. SUBJECT TERMS			15. NUMBER OF PAGES	
thunderstorms, moisture, lift, instability, Florida			93	
			16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT		18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified		Unclassified	Unclassified	UL